Title
Neurocomputational mechanisms of timing, temporal context, and working memory

Author
Seay, Michael James

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Neurocomputational mechanisms of timing, temporal context, and working memory

A dissertation submitted in partial satisfaction of the

requirement for the degree Doctor of Philosophy

in Psychology

by

Michael James Seay

2022

ABSTRACT OF THE DISSERTATION

Neurocomputational mechanisms of timing, temporal context, and working memory

by

Michael James Seay

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2022

Professor Dean Buonomano, Chair

Humans effortlessly parse continuous experience based on its temporal structure, allowing us to recognize speech, detect regularity in sequences of events, and predict when things will happen. Yet it remains poorly understood how the nervous system accomplishes this multifaceted effort. In the current dissertation, I first review organisms' abilities to sense timing on the scale of tens to hundreds of milliseconds, as well as evidence of sensory neurons that respond selectively based on temporal features or respond differently based on recent temporal context. I propose that neuronal selectivity to timing results from time-varying neural and synaptic properties, most notably short-term synaptic plasticity (STP), and I review supporting evidence. Next, I present a computational model that explains why different sensory neurons show different patterns of sensitivity to temporal context. Like real neurons observed in mice, model neuron responses either decrease, remain stable, or increase over the course of

repeated stimulation on short timescales, an effect that relies on the model's usage of experimentally-observed STP at synapses with two distinct types of inhibitory interneurons. I test and confirm model predictions by analyzing the responses of mouse auditory neurons to trains of repeated pure tones. Subsequently, I shift my focus toward a potential mechanism of internally-generated timing on the circa-second scale: persistent activity states. I build on recent computational work by showing that counter-intuitive "cross-homeostatic" plasticity rules are able to configure neural networks to exhibit stable persistent activity states in a large, sparsely-connected spiking model. Importantly, I show that when cross-homeostatic plasticity operates using only local signals, it fails unless counterbalanced by classical homeostatic plasticity. Finally, I test the idea that timing is computationally linked with working memory by performing behavioral experiments in humans. To do so, I employ two tasks that have the same stimulus structure but differ in whether timing or working memory is required to respond correctly. I find that in each task participants learn about and use the other task-irrelevant component, which is consistent with the hypothesis that in some cases working memory and timing information are multiplexed in a time-varying format because of the importance of predicting when working memory will be used.

The dissertation of Michael James Seay is approved.

H. Tad Blair

Dario L. Ringach

Jesse A. Rissman

Dean Buonomano, Committee Chair

University of California, Los Angeles

2022

This dissertation is dedicated

to the memories of my brother, William Reynolds Seay, III,

and my mother, Catherine Lisa Helm Seay.

Table of Contents

List of Figures

List of Tables

List of Acronyms

Abbreviation	Meaning
A1	primary auditory cortex
BOS	bird's own song
dDMS	differentially delayed match-to-sample
EPSP	excitatory postsynaptic potential
FR	firing rate
IES	inverse efficiency score
IPSP	inhibitory postsynaptic potential
ISA	interval stimulus association
ISN	inhibition-stabilized network
MACW	mean absolute change in weights
MSE	mean squared error
PSTH	post-stimulus time histogram
PV	Parvalbumin, a type of inhibitory interneuron
PV-Cre	an animal in which a Cre-dependent virus is expressed in PV cells
Pyr	Pyramidal cell, the primary excitatory cell type in cortex
RT	reaction time
STP	short-term synaptic plasticity
SSA	stimulus-specific adaptation
SST	Somatostatin, a type of inhibitory interneuron
SST-Cre	an animal in which a Cre-dependent virus is expressed in SST cells
WM	working memory

Acknowledgments

I thank my advisor Dr. Dean Buonomano for taking me on as a student and nurturing my growth. From the moment I began in the lab, he pushed me outside of my comfort zone while trusting in my ability to learn to perform research that I never thought I could. And when my world was turned upside down by unexpected events, he was infinitely understanding and supportive. For this I will be forever grateful.

I thank my research family in the Buonomano Lab: Drs. Nick Hardy, Vishwa Goudar, Helen Motanis, Shanglin Zhou, and Saray Soldado-Magraner, as well as Juan Luis Romero Sosa, Ben Liu, and Ashita Tanwar. Your hard work, kindness in teaching me, and thoughtful discussions enabled this dissertation.

I also thank my collaborators Dr. Maria Geffen at the University of Pennsylvania and Dr. Ryan Natan at UC Berkeley, whose fascinating experimental data sparked a dialog with my model that resulted in Chapter 2. And I thank Drs. Sotiris Masmanidis, Peyman Golshani, and Jiannis Taxidis at UCLA, whose experiments inspired Chapter 4.

I thank the institutions at UCLA that supported my academic journey, including the Departments of Psychology & Neurobiology and the Integrative Center for Learning & Memory. Your ferocious interdisciplinary pursuit of scientific truth inspired me throughout my program of study.

Finally, I thank the Edwin W. Pauley Fellowship program and the Graduate Summer Research Mentorship program for providing financial support.

Vita

Education

2018 M.A., Psychology, University of California, Los Angeles, CA

2013 M.S., Neuroscience, Tulane University, New Orleans, LA

2012 B.S., Neuroscience, Tulane University, New Orleans, LA

Research Experience

2017-2022 Buonomano Lab at UCLA (Advisor: Dr. Dean Buonomano)

2015-2017 Henri Begleiter Neurodynamics Laboratory at SUNY Downstate Medical Center, New York, NY (Advisor: Dr. Bernice Porjesz)

2010-2015 Tulane Cognitive Neuroscience Laboratory at Tulane University, New Orleans, LA (Advisor: Dr. Edward Golob)

Publications

Seay MJ, Natan RG, Geffen MN, Buonomano DV (2020). Differential Short-Term Plasticity of PV and SST Neurons Accounts for Adaptation and Facilitation of Cortical Neurons to Auditory Tones. *Journal of Neuroscience.*

Motanis H*, **Seay MJ***, Buonomano DV (2018). Short-Term Synaptic Plasticity as a Mechanism for Sensory Timing. *Trends in Neurosciences*. *contributed equally

Mock JR, **Seay MJ**, Charney DR, Holmes JL, Golob EJ (2015). Rapid cortical dynamics associated with auditory spatial attention gradients. *Frontiers in Neuroscience*.

Chapter 1: Neurocomputational mechanisms for sensory timing and temporal contextual processing

Based on: Motanis H*, Seay MJ*, Buonomano DV. *Short-Term Synaptic Plasticity as a Mechanism for Sensory Timing*, 2018. Trends in Neurosciences. 41(10): p. 701-711.

Abstract

The ability to detect time intervals and temporal patterns is critical to some of the most fundamental computations the brain performs, including the ability to communicate and appraise a dynamically changing environment. Many of these computations take place on the scale of tens to hundreds of milliseconds. Electrophysiological evidence shows that sensory responses of some neurons are selective to the temporal features of driving stimuli including duration, interval, rate, or order. More generally neural responses to simple stimuli can be modulated based on the recent temporal context within a sequence of stimuli. Because the time constants of many time-varying neural and synaptic properties, including short-term synaptic plasticity (STP), are also in the range of tens to hundreds of milliseconds, they are strong candidate mechanisms underlying selectivity to temporal features and temporal context. Neurophysiological studies indicate that STP is indeed one of the mechanisms that contributes to temporal selectivity, and computational models demonstrate that neurons embedded in local microcircuits can exhibit temporal selectivity if their synapses undergo STP. Converging

evidence suggests that some forms of temporal selectivity emerge from dynamic changes in the balance of excitation and inhibition imposed by STP.

Chapter 1: Neurocomputational mechanisms for sensory timing and temporal contextual processing

> *"How would you describe time?"*
> *The Big Business Man smiled. "Time," he said, "is what keeps everything from happening at once."*
> *"Very clever," said the Chemist, laughing.*
>
> Ray Cummings, "The Girl in the Golden Atom," 1919

1.1 The challenge of time

Natural events proceed over time. When one listens to the radio over the course of several minutes or observes the growth of a plant over the course of a week, one engages stimuli that unfold and change from moment to moment. Yet, from the point of view of an organism, sequential changes in external stimuli are embedded in a continuously evolving stream of sensory input. In this sense, the temporal structure of events presents both a *challenge* and a valuable source of *information*. For example, the sub-millisecond difference between a sound's arrival to the left and right ear can be used to localize an approaching predator, and the daily change in light that accompanies the movement of the sun can be used to calibrate when to gather food. But to capitalize on the information contained in the temporal structure of external inputs, animals must have adapted biological mechanisms to do so.

As the previous examples illustrate, animals have evolved mechanisms to tell time on scales spanning more than ten orders of magnitude[1], but it is on the scale of tens-to-hundreds of milliseconds that our ability to tell time and extract temporal information is at its most sophisticated. Within this range, we are not only able to identify simple temporal intervals but extract higher-order temporal patterns. Speech

comprehension, for example, requires extraction of a hierarchy of temporal information: from the voice-onset time of syllables (which contributes to the /ba/ versus /pa/ distinction, for instance), to phrasal boundaries, to prosody[2, 3]. Indeed, speech can be recognized even when spectral information is impoverished but temporal structure is preserved, meaning that the temporal envelope alone provides a significant amount of information for speech recognition[4, 5].

Importantly, even on the subsecond scale, timing is not a unitary problem, but encompasses a range of interrelated problems necessary for sensorimotor processing, learning, and cognition[6-8]. In the current chapter, I propose that temporal selectivity is an intrinsic property of neural circuits that relies on time-varying synaptic and neuronal properties, and I highlight short-term synaptic plasticity as one of the key mechanisms in the emergence of temporal selectivity. In the following chapters, I first focus on mechanisms of sensory temporal selectivity at the tens-to-hundreds of milliseconds scale, but then move on to the circa-second scale by focusing on mechanisms of internally-generated timing and its relationship with other cognitive phenomena. Specifically, I focus on the neurocomputational mechanisms underlying three distinct empirical phenomena: 1) short-term sensory adaptation in mouse auditory cortex, 2) the emergence of stable persistent activity states in excitatory-inhibitory networks, and 3) the relationship between interval timing and working memory.

1.2 Evolved timing behaviors

To approach the goal of understanding how the nervous system extracts information from the temporal structure of sensory inputs, we should first look to the natural world. Scientists and naturalists have observed a plethora of naturally-occurring

timing behaviors on the scale of tens-to-hundreds of milliseconds. These behaviors provide an opportunity to understand the ethological role of sensory timing and an empirical dataset for neuroscientists to study evolved solutions for overcoming the challenge of timing.

In the auditory modality, timing information is particularly prominent. Many species of animals use acoustic signals for communication, courtship, territoriality, social affiliation, and conspecific recognition[9, 10]. Acoustic communication relies not only on spectral signatures (e.g., pitch) but on temporal features such as interval, duration, rate, and overall temporal structure. For example, cicadas and grasshoppers produce songs that consist of rhythmic sequences of sound pulses, and they use the temporal pattern of acoustical pulses to recognize conspecifics[11, 12]. In courtship, female crickets exhibit phonotaxis, a behavior characterized by walking or flying toward singing males, and phonotaxis is strongest at pulse durations and intervals that are within the range of the male calling song parameters[13, 14]. Similarly, some frog species use the duration and interval of acoustic pulses to differentiate between conspecific and heterospecific calls[12]. Changing the interval between a single pair of pulses in a male frog’s ten-pulse mating call significantly decreases the percentage of female frogs showing attraction[12].

Timing behaviors also extend beyond the auditory modality, and they do not always serve a communicative function. A non-auditory example of sensory timing comes from weakly electric mormyrid fish that utilize intervals between successive electric organ discharges[15]. Mormyrids produce patterns of eight to twelve electric pulses called scallops that identify the sender and play a role in courtship and

aggression behaviors[16]. And one of the best studied examples of sensory timing for a non-communicative function is echolocation in bats. Specifically, bats use the intervals between emitted acoustic pulses and their echo to navigate their environments and determine the position of potential prey[17].

Many forms of temporal processing rely on experience, highlighting the role of learning in sensory timing. Rodents, for example, can be trained to make temporal judgments as to whether intervals are short or long relative to each other[18, 19]. And humans are capable of robust temporal perceptual learning, which is generally reported to be interval-specific. For example, repeated interval discrimination of an auditory interval of 100 ms leads to improved discrimination around this interval, but not to shorter or longer intervals[20, 21].

The above examples establish that animals extract information from the temporal features of sensory events. Thus, they must possess some internal mechanisms for processing temporal features of sensory input streams. In the next section, I will review evidence of neurons that respond selectively to features such as interval and duration—i.e., temporally-selective neurons.

1.3 Temporal feature selectivity in sensory neurons

Neurons that are tuned to temporal features such as interval, duration, pulse rate, and temporal structure of vocalizations have been reported in multiple species [22-26] (**Fig. 1.1**). Timing-sensitive neurons have often been discovered through inquiry into the neural correlates of behaviors discussed in the previous section. For example, phonotactic tuning in female crickets to parametrically manipulated calling songs is mirrored in the tuning of their sensory neurons' firing rates to the same stimuli[14]. And

neurons in the midbrain of the weakly electric mormyrid fish have been found to be tuned to pulse rate, spiking with low probability for pulse rates of 10 or 100 Hz but with high probability for a rate of 20 Hz [27-29]. (**Fig. 1.1A**). *In vivo* intracellular recordings have shown that these neurons' pulse rate-sensitivity allows them to be sensitive to the precise temporal structure of scallops, discriminating natural scallops from time-reversed, randomized, and jittered sequences[27, 29].

In addition to interval and rate tuning, many animals also possess neurons that are tuned to the duration of stimuli. Elegant examples of duration-tuned neurons come from studies in the brainstem of echolocating bats. Inferior colliculus neurons in echolocating bats are tuned to the duration of the high-frequency pulses that they emit [30-32], and the preferred durations match the range of durations used in echolocation signals[32-34]. More generally, duration-tuned neurons have been found in the central auditory systems of frogs[35-37], rodents[38, 39], chinchillas[40], and cats[41]. Duration-sensitive neurons have also been observed in the visual modality: off responses recorded from neurons in the cat visual cortex can be tuned to the duration of a stationary bar of light[42] (**Fig. 1.1B**). The presence of duration-tuned neurons across species and sensory modalities suggests that duration selectivity is a general property of sensory systems.

An important question pertaining to the temporally-tuned neuronal responses mentioned above is whether they reflect innate hardwired circuits, or rather emerge in an experience-dependent manner as a result of learning and plasticity. It seems likely that in some animals temporal selectivity reflects, at least in part, hardwired circuits. But

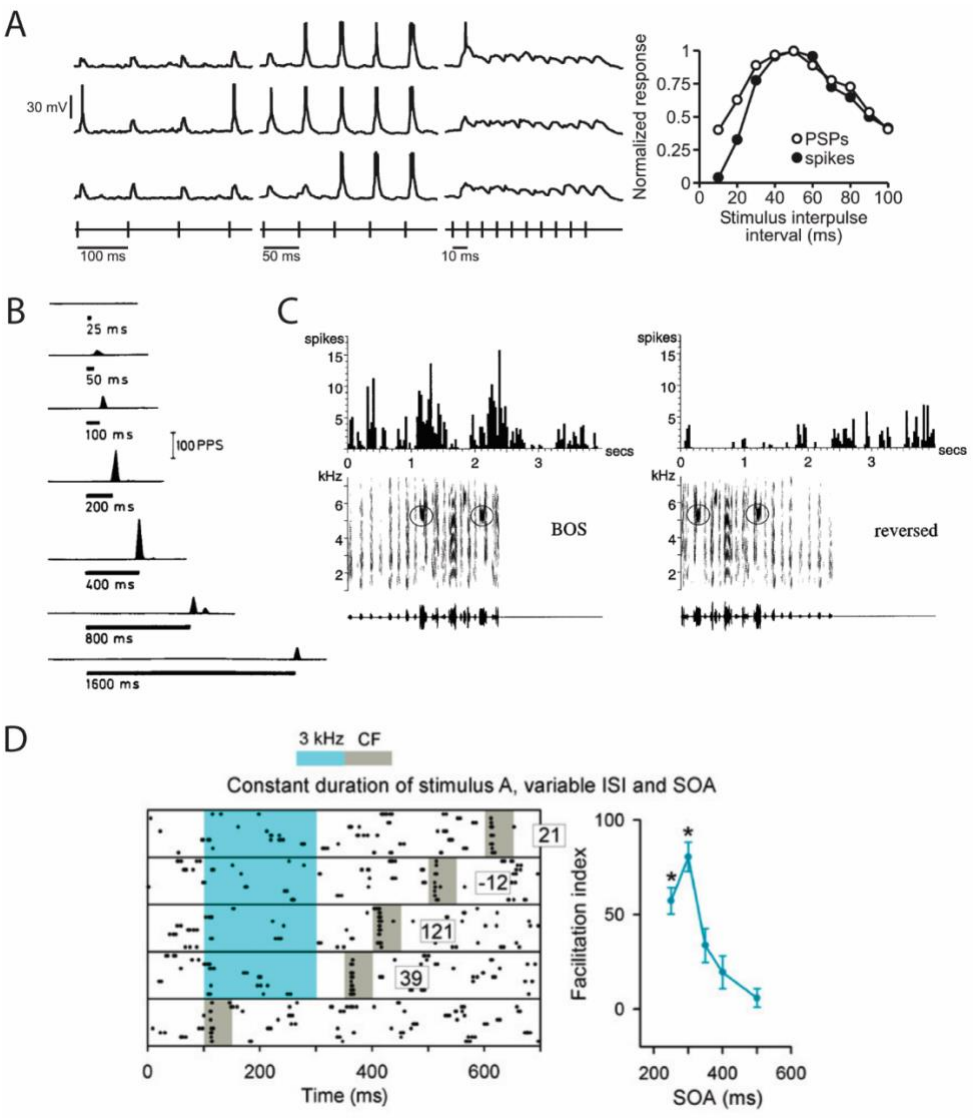

Figure 1.1 Temporal-selective neurons across species and modalities.
A. Interval-sensitive neuron in the midbrain of an electric fish. Voltage traces were recorded during the presentation of trains of electrical pulses at intervals of 100 ms (left), 50 ms (middle) or 10 ms (right). Rows represent three separate repetitions of each train. At right, the amplitude of the postsynaptic potentials (PSPs) and number of spikes elicited by trains at each interval is plotted, demonstrating that the recorded neuron was tuned to pulses delivered at intervals of 50 ms. Reproduced from [27]. **B**. Duration-tuned neuron in cat visual cortex. The recorded neuron produced off responses to a static bar of different durations, with maximal response to a duration of 400 ms. Of 174 neurons, ~30% responded differentially to duration, and 3% showed sharp duration tuning curves. Reproduced from [42]. **C**. Complex spectrotemporal pattern-selectivity of a single neuron in the lateral magnocellular nucleus of the anterior neostriatum in an adult zebra finch. Response to the bird's own song (BOS) in forward (left) and reversed (right) order. Below each spike plot, the song's spectrogram and waveform are shown. Adapted from [43]. **D**. Interval-sensitive neuron in the auditory cortex of a rat. Spike raster (left) in response to five different stimuli, each composed of a 200-ms 3-kHz tone followed by a 50-ms 7-kHz tone (the recorded neuron's characteristic frequency; CF) with different stimulus-onset asynchrony (SOA). Numbers represent the facilitation index. Graph (right) shows the average interval-tuning curve. The rat had been trained to detect an inter-tone onset interval of 300 ms (middle row on left). Reproduced from [44].

in other cases, it is clear that temporal neuronal selectivity emerges in an experience-dependent fashion (and as mentioned previously, many animals can learn to discriminate intervals and durations). One of the clearest examples of experience-dependent acquisition of complex stimulus selectivity comes from songbirds. Like speech learning, song acquisition occurs early in a songbird's life, and is critically dependent on auditory experience and feedback[45]. Neurons in multiple areas of the adult male finch brain are strongly selective for both spectral and temporal properties of birdsong; they respond more robustly to the bird's own song (BOS) than to songs of conspecific individuals, and they respond less well to the BOS if it is played in reverse[43, 46-48] (**Fig. 1.1C**).

Such experience-dependent emergence of temporally selective neurons has also been observed in mammals exposed to or trained on stimuli defined by interval, duration, or order of the underlying tones[44, 49-51]. For example, in one study rats were trained on a go/no-go task with a target stimulus composed of a 3 kHz tone followed by a 7 kHz tone with an inter-onset interval of 300 ms[44]. Recordings in primary auditory cortex revealed a substantial number of neurons that responded optimally at this interval, indicating that learning was accompanied by the formation of auditory neurons that were tuned to the spectrotemporal features of the target stimulus (**Fig. 1.1D**).

Tuning to *spatial* features is among the most widely studied aspects of sensory systems—ranging from selectivity to specific orientations of visual lines to selectivity to the frequency of tones (which I consider "spatial" because of the tonotopic organization of the cochlea). The studies discussed above suggest that selectivity to temporal

features—e.g., duration, interval, rate, and order of sensory events—is also prevalent among sensory neurons.

1.4 Temporal context

Although the previous examples illustrate that nervous systems are capable of representing specific temporal features in specialized or learned stimuli, they do not address general capacities for parsing sequences of sensory events based on their temporal structure. It is well-known that the spatial arrangement of a pattern of visual stimuli will influence its perception (e.g., if the spacing between columns in a square grid of nine dots is increased, it will be perceived as three groups of three). Similarly, the *temporal* arrangement of stimuli influences the way in which an input sequence is perceived. Temporal features such as interval and duration influence perceptual grouping vs. separation, while repetition and regularity in temporal structure allows differentiation of redundant vs. novel stimuli[52-55]. This general property can be referred to as temporal context, and the ability to perceive the same stimulus or set of stimuli differently depending on temporal context can be considered a form of temporal processing, in that it is dependent on the temporal structure of the input sequence.

One ubiquitous example of how temporal context of sensory stimuli shapes neuronal responses is the phenomenon of short-term sensory adaptation (**Fig. 1.2A**). Across modalities, sensory cortical neurons decrease their evoked firing rates to identical stimuli when they are repeated on short timescales – i.e., with interstimulus intervals up to several hundred milliseconds[56-65]**.** A closely related phenomenon that takes place with extensive repetition over longer timescales is stimulus-specific adaptation (SSA, **Fig. 1.2B**). In SSA, neurons selectively reduce their responses to a

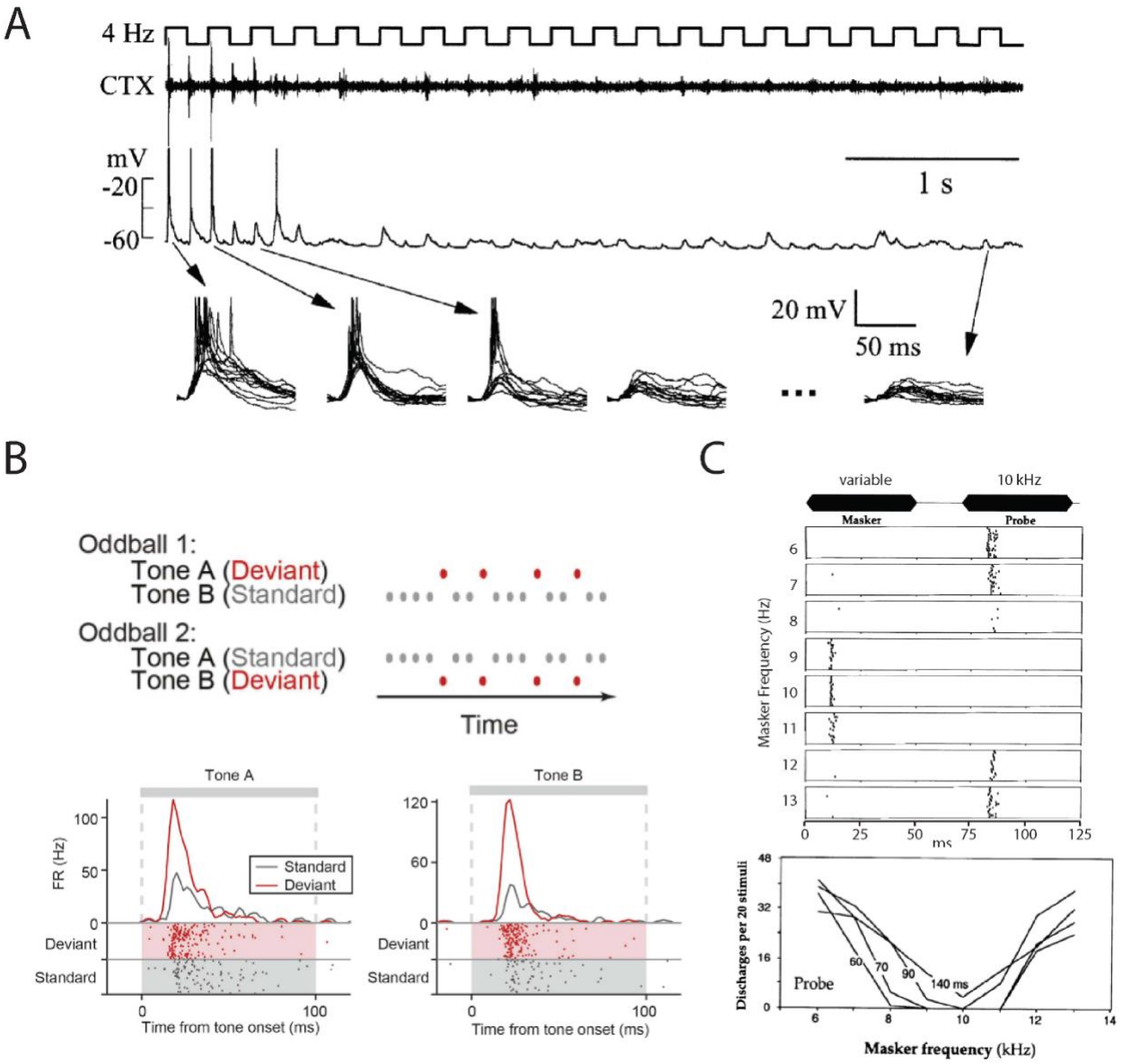

Figure 1.2 Examples of temporal contextual modulation.
A. Short-term sensory adaptation. Recordings from rat somatosensory cortex during repeated 4 Hz stimulation of the same whisker. The upper trace of square pulses indicates the stimulus train. Middle and lower traces are extracellular and intracellular recordings, respectively. Multiple overlaid trial responses of the first four and the last intracellular responses are shown below at an expanded time scale. Reproduced from [62]. **B**. Stimulus-specific adaptation (SSA). A 2.5 Hz sequence of tone pips at two possible frequencies. In two separate sequences, one frequency was presented 90% of the time (standard), while the other was presented 10% of the time (deviant). In the other sequence, the roles were reversed. Lower plots show the response of a single neuron recorded with an extracellular probe in mouse auditory cortex to tone A (left) and tone B (right) depending on its probability. Reproduced from [66]. **C**. Forward suppression. Maskers of variable frequency preceded a 10 kHz probe tone. Each upper raster plot displays the spiking response of a single neuron in cat auditory cortex to 20 repetitions of a two-tone sequence with maskers from 6-13 kHz. The lower overlaid line plot shows the total number of spikes to the 10 kHz probe tone as a function of the masker frequency for various delays. Reproduced from [67].

“standard” tone that comprises 90% of the tones in a continuous train of tones while responding robustly to rare “oddball” tones presented at a different tone frequency [54, 66]. For many neurons, this relationship persists even when the roles of standard and oddball are reversed, indicating that SSA is not a result of frequency tuning but of the temporal structure of the sequence[66].

More generally, the response of auditory cortical neurons to a probe tone can be suppressed to different extents depending on the frequency of a single preceding masking tone (**Fig. 1.2C**), a phenomenon referred to as forward suppression[67-71]. For most units, the response to the probe depending on the masker frequency takes a “U” shape in which suppression deepens as the frequency of the masker is made more similar to the probe. However, for some units, presenting maskers that are distant in frequency actually causes an enhancement of the response to the probe – i.e., “forward facilitation” [70-72] – which indicates that the temporal contextual modulatory effect is complex and diverse in the population response.

The possibility of response enhancement to arbitrary spatiotemporal patterns of stimuli opens up the way for general sequence sensitivity. Auditory cortical neurons that exhibit enhanced responses to sequences of simple stimuli such as pure tones, with no training or reinforcement, have been characterized in cats and monkeys[73-75]. Given that pure tones exemplify the basic components of auditory objects (i.e., fundamental frequencies), these observed sequence-selective neurons provide the representational substrate necessary for fundamental capacities such as order discrimination. And even in the case of the repeated presentation of an identical tone at moderate rates (e.g. every 400 ms), progressive increases in the evoked firing rate of single auditory cortical

neurons over the course of repetition have been observed in a minority of units in mouse auditory cortex[76]. Thus, the possibility of both suppression and facilitation in firing rates due to temporal context – and the mixture of both in the population response – may provide representational flexibility when sequence information could be either redundant or critical, depending on the behavioral circumstances.

Crucially, the temporal contextual functions reviewed above only operate when the interval between consecutive stimuli is within the tens-to-hundreds of millisecond range, indicating that the timing of temporal contextual interactions are calibrated to the average rate of environmental stimuli. For humans, this matches well with the rates of stimuli such as speech and music, and this explains why speech and music that is too fast or slow degrades in meaning.

1.5 Neurocomputational mechanisms for sensory timing and temporal context

The breadth of examples across species and modalities suggests that neural sensitivity to temporal features on the order of tens-to-hundreds of milliseconds reflects a general computation within sensory circuits. One hypothesis is that temporal tuning is an intrinsic property of local neural circuits that relies on time-varying synaptic and neuronal properties. Neurons and synapses possess an abundance of functional properties with time constants on the scale of tens-to-hundreds of milliseconds that have been proposed to contribute to sensory timing, including ionotropic and metabotropic receptors[77], ion channels[25, 78, 79], and most notably short-term synaptic plasticity (STP)[28, 80-84]. Below I focus on the contribution of STP to sensory timing but emphasize that other neural properties have also been implicated, perhaps

most notably dynamic changes in the excitation/inhibition balance and rebound excitation [28, 85-87].

1.5.1 Short-term synaptic plasticity (STP)

STP refers to use-dependent changes in the strength of synaptic connections that take place on time scales of tens to hundreds of milliseconds[88]. At a synapse exhibiting STP, trains of presynaptic spikes that occur within a short timespan cause progressively smaller or larger postsynaptic potentials (**Fig. 1.3**). These two opposing forms of STP are referred to as short-term depression (or paired-pulse depression) and short-term facilitation (or paired-pulse facilitation) respectively. These two broad forms of STP, however, can interact to form more complex temporal profiles[89]. Short-term depression results primarily from exhaustion of readily-releasable vesicles in the presynaptic terminal. Short-term facilitation, although less precisely understood, involves an increase in probability of vesicle release due to residual presynaptic calcium, which may involve the activation of specialized presynaptic calcium sensors[88, 90].

STP is remarkably diverse across neuron types[91-94], cortical layers[95], brain regions[96, 97], and can be modulated by development[98-100], sensory experience[101], brain state[102], and by neuromodulation[103]. Despite this richness and diversity, some general principles have emerged. For example, although STP is generally attributed to presynaptic mechanisms, the nature of STP of excitatory synapses primarily depends on the postsynaptic cell type, while the nature of STP of inhibitory synapses primarily depends on the presynaptic cell type[89, 92, 104]. For excitatory-to-inhibitory synapses, excitatory postsynaptic potentials (EPSPs) onto fast-

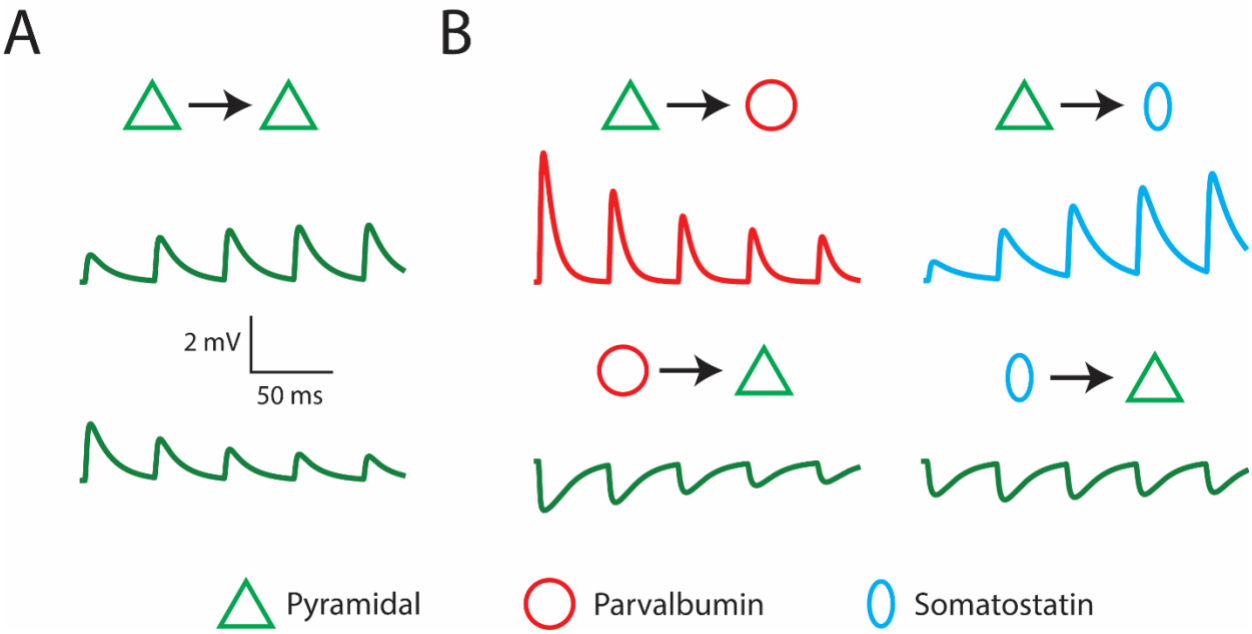

Figure 1.3 Diversity of short-term synaptic plasticity (STP) at cortical synapses.
Simulations of synaptic transmission with STP based on real whole cell recordings.Traces indicate voltage of the postsynaptic cell as the presynaptic cell fires a train of five action potentials at 20 Hz. Diagrams above each trace indicate the identity of pre- and postsynaptic neurons, including pyramidal (green triangle), Parvalbumin-positive inhibitory (red circle), and Somatostatin-positive inhibitory (cyan oval) cells. **A.** Facilitating (top) and depressing (bottom) inter-pyramidal synapses. Based on recordings performed for [98]. **B**. Excitatory-to-inhibitory (top row) and inhibitory-to-excitatory (bottom row) synapses. Based on recordings performed for [105].

spiking inhibitory parvalbumin-positive interneurons generally undergo moderate depression[91, 106, 107], whereas EPSPs onto low-threshold-spiking somatostatin-positive inhibitory interneurons generally exhibit strong facilitation[107, 108] (**Fig. 1.3B**).

1.5.2 The role of STP in temporal selectivity

Even though STP is observed across virtually all synapses, there is no consensus as to its computational function[109, 110]. STP has been hypothesized to enable dynamic gain control[111, 112] as well as sensory adaptation and sensitization. More generally, it is recognized that STP can implement temporal filters[80-82, 113, 114]—that is, STP transforms temporal patterns of presynaptic spikes into different postsynaptic patterns depending on the STP characteristics of the activated synapses.

The ability to implement temporal filters at various timescales means that, at least theoretically, STP has the potential to underlie temporal selectivity in neurons[81]. For example, a simulation of a simple circuit composed of integrate-and-fire units demonstrates how STP can be used to generate interval selectivity (**Fig. 1.4**). In this simulation, an input unit forms facilitating synapses onto both an excitatory (Exc) and an inhibitory (Inh) unit that provides feedforward inhibition onto the Exc unit (**Fig. 1.4A**). As the Input unit generates spike pairs separated by intervals of 50, 100, or 200 ms in separate trials, the resulting EPSPs facilitate to different degrees (**Fig. 1.4B**). With appropriate tuning of synaptic weights, this simple circuit can function as an interval

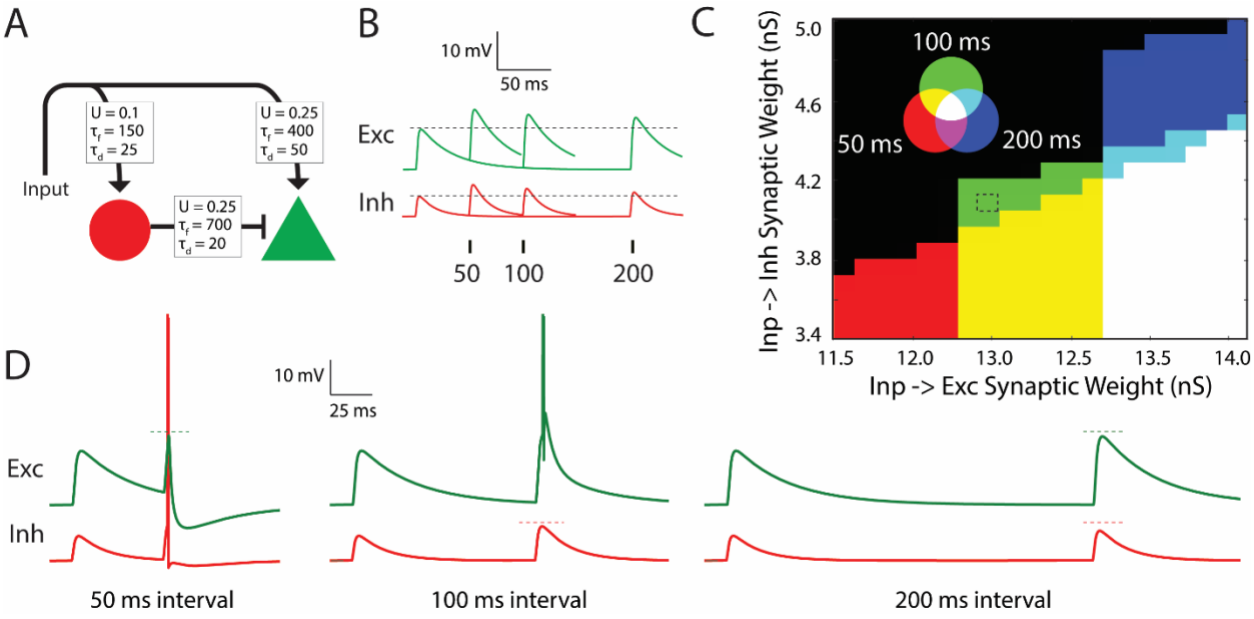

Figure 1.4 Interval selectivity simulated in a simple circuit with STP.
A. Circuit composed of a single inhibitory (red circle) and excitatory (green triangle) neuron. Each synapse is endowed with STP [115]. Parameters used to simulate STP are shown, including the baseline release probability (U) and the time constants of facilitation (τ_f) and depression (τ_d). **B.** Excitatory and inhibitory responses to separate input spike intervals of 50, 100, and 200 ms are overlaid, illustrating short-term facilitation. **C.** The weights of the Input→Exc synapse (x-axis) and Input →Inh synapse (y-axis) are varied, and the interval selectivity at each point is examined. The color in each region indicates the responsiveness of the Exc unit to one or more of the input intervals. Black areas represent regimes in which the Exc unit fired to the first pulse only, or did not fire at all. **D.** Voltage traces of the Exc and Inh units during simulations with synaptic weights that resulted in a 100 ms interval detector (dotted rectangle in C). Based on [116].

detector in which the Exc unit is the readout (**Fig. 1.4C**). For example, there is a range of synaptic weights at the Input→Exc and Input→Inh connections in which the Exc unit fires exclusively to the 100 ms interval (**Fig. 1.4D**). This selectivity emerges because, for the 200 ms interval, short-term facilitation at the Input→Exc synapse has had more time to decay and thus the Exc unit's EPSP is subthreshold, while for the 50 ms interval, short-term facilitation at the Input→Inh synapse is strong enough to drive the Inh unit to spike, thus vetoing what would be a suprathreshold EPSP in the Exc unit.

Over the past decade converging experimental evidence has provided support for hypotheses suggesting that STP contributes to temporal selectivity. For example, STP appears to underlie temporal selectivity in the anuran auditory system[117], in which two broad classes of temporally-selective neurons have been identified. One class consists of short-interval cells that respond best when presented with an optimal number of pulses presented at faster rates[118]. Short-interval cells respond to consecutive inputs with EPSPs followed by large, slow inhibitory postsynaptic potentials (IPSPs). Selectivity appears to result from an enhancement of EPSPs elicited by repeated pulses—that is, a progressive enhancement in EPSP magnitude is eventually able to overcome the strong but stable inhibitory response to each pulse. Importantly, enhancement of excitation is optimal for certain pulse rates[119]. A second class of temporally-selective cells in anuran auditory systems responds well only to slow pulse rates but fails to respond to fast pulse rates. Electrophysiological experiments suggest that the low-pass properties of these neurons resulted from cancellation of temporally-offset excitatory and inhibitory synaptic inputs at fast pulse rates, together with short-term synaptic depression at high stimulation rates[120].

Additional experimental work regarding the mechanistic involvement of STP in pulse rate selectivity comes from whole-cell recordings of neurons in weakly electric mormyrid fish[27-29]. By estimating synaptic conductances during temporally-selective responses, Baker and colleagues determined that both excitatory and inhibitory conductances exhibited short-term depression. However, for high-pass neurons (neurons tuned to faster pulse rates), inhibitory conductances depressed more strongly than excitatory conductances, while for most low-pass neurons excitation depressed more strongly[28]. Analytically reconstructing cellular responses while excluding short-term depression led to drastically reduced diversity in interval tuning[28].

1.5.3 Neural and synaptic mechanisms of temporal contextual modulation

STP has long been postulated to be a primary mechanism underlying the firing rate suppression that accompanies short-term sensory adaptation and forward suppression[62, 69]. Specifically, it has been proposed that depression at thalamocortical synapses causes the progressive decrease in firing evoked by identical repeated tones[62]. More recently, optogenetic experiments in mice have additionally implicated the involvement of inhibitory PV and SST interneurons[66, 71, 76], suggesting that intra-cortical inhibition, which is modulated by STP at excitatory-to-inhibitory synapses, plays a role in short-term contextual modulation. After all, excitatory synapses onto SST interneurons should facilitate with repetition, and thus SST cells should provide a progressively larger inhibitory force in sensory circuits over the course of repetition. Optogenetic experiments have supported this idea. One experiment showed that optogenetically inactivating SST interneurons during the eighth but not the first tone in a sequence of identical repetitions significantly increased the tone-evoked

firing rate of excitatory units[76]. In an oddball paradigm designed to demonstrate stimulus-specific adaptation, inactivating SST interneurons increased the response of excitatory units to the standard but not the oddball tone, an effect that emerged as standards were repeated multiple times[66]. Finally, another experiment showed that optogenetically inactivating SST interneurons during forward suppression decreased the number of excitatory units exhibiting suppression as well as the degree of suppression[71].

Mechanistic understanding of short-term facilitation phenomena is less well defined. Nevertheless, optogenetic inactivation experiments have also provided some support for the idea that, opposite to SST, the depression at excitatory-to-PV synapses may contribute to the stability or enhancement of responses that is often observed among excitatory units[71, 76]. One study found that inactivating PV units during forward suppression increased the number of suppressing units as well as the degree of suppression[71]. This finding provides support for the idea that the depression of feedforward PV inhibition may help to stabilize or enable enhancement of excitatory responses to identical stimuli.

Aside from experimental studies, multiple computational models have recently used STP to successfully account for short-term sensory adaptation, stimulus-specific adaptation, and forward suppression[66, 70, 121, 122]. In chapter two of the current work, I offer a reduced spiking model of feedforward inhibition in sensory cortex that addresses both short-term sensory adaptation and facilitation to identically repeated stimuli [76, 122].

1.5.4 Network models of temporal pattern selectivity based on STP

The theoretical and experimental evidence discussed above indicate that STP plays a role in temporal filtering and the formation of temporally selective neurons. Indeed, as shown in **Fig. 1.4**, it is relatively straightforward to create interval selective neurons in disynaptic circuits that exhibit short-term facilitation. However, in this example, interval selectivity relies on the careful tuning of synaptic weights and STP. Far more general models of cortical computation referred to as state-dependent network models or liquid state machines[80, 81, 123, 124] propose that STP provides a rich mechanism to endow cortical networks with the ability to decode the spatiotemporal structure of stimuli. Specifically, STP functions as a memory of what happened within the past few hundred milliseconds. Consider the case of two identical tones arriving in the auditory cortex 100 ms apart during an interval discrimination task. Even if we assume the second tone activates the same pattern of thalamocortical inputs into the cortex as the first tone, it will arrive in a different cortical state, where some synapses will be depressed and others facilitated. Thus, the same tone should have a different net effect on the circuit, depending on the recent input history. While some neurons will be activated by both events, others are likely to be activated by one or the other, and these neurons can provide information about the length of the interval or the order of events.

In these models, STP (and other time-varying properties) provides a memory buffer that ensures that each event is encoded in the context of the previous events. Thus if two tones A and B are presented 100 ms apart, the response to B does not simply encode the stimulus B, but ‘B preceded by A’. This view predicts that it should be possible to decode previous stimuli based on the population response to the current

stimulus. This prediction has been confirmed, by showing that in the visual cortex, when a pair of images is sequentially presented it is possible to determine the first image based on the response to the second[125]. Another prediction is that interval discrimination should be impaired by preceding stimuli, and indeed psychophysical experiments show that simply presenting two intervals to be judged close together in time impairs interval discrimination[126, 127]. While these results are consistent with the role of STP in establishing the state-dependence of the local network (the memory buffer), it remains to be determined whether STP is indeed one of the mechanisms underlying these results. Some support to this possibility comes from computer simulations, which have established that randomly connected recurrent neural networks endowed with STP are intrinsically capable of discriminating simple intervals[80, 81, 123, 126, 128]. Furthermore, the presence of STP in such networks enhances their ability to discriminate complex temporal patterns such as speech[81, 129, 130].

1.6 Summary and overview of following chapters

Time presents a challenge. For many organisms, the way in which their nervous systems have evolved to overcome this challenge fundamentally defines their mental world, allowing them to communicate and make sense of the sequential structure present in sensory streams. Across species and modalities, neuroscientists have observed sensory neurons that are sensitive to temporal features such as interval, duration, and overall spatiotemporal structure. Research thus far suggests that time-varying synaptic and neural properties, most notably STP, not only support temporal selectivity but also provide more general properties such as temporal context sensitivity and state dependence. In the following chapters, I first focus on neurocomputational

mechanisms of temporal contextual processing at the tens-to-hundreds of milliseconds scale, but then I move on to the circa-second scale by focusing on mechanisms of internally-generated timing and its relationship with other cognitive phenomena.

First, chapter 2 tests the hypothesis that STP is a key mechanism underlying temporal contextual modulation using a unique experimental dataset of short-term sensory adaptation in mouse auditory cortex[76]. I present a reduced spiking model that accounts for the diversity of observed adaptation “profiles” with differential STP of excitatory synapses onto PV and SST interneurons. I confirm the model predictions with novel analyses of the experimental data, showing that: 1) Cortical neurons with steady firing rates across repeated stimuli exhibit shorter firing latencies than those with adapting or facilitating profiles, and 2) optogenetically inactivating PV interneurons during the first stimulus in a train of repeated stimuli causes a decreased excitatory response to the second stimulus 400 ms later. The results provide evidence that shifts in firing latency caused by STP are critical for short-term temporal contextual modulation.

Next, chapter 3 focuses on the long-term plasticity rules that configure neural circuits to support persistent activity states, which are mechanistically relevant to internally-generated timing phenomena. Recent computational work in my lab has shown that a family of “cross-homeostatic” plasticity rules is capable of robustly configuring synaptic weights to support stable persistent activity states in fully-connected firing rate models. I reinforce and build on this work by demonstrating the robustness of the cross-homeostatic plasticity rules in a large sparsely-connected spiking model. Further, due to its sparse connectivity, my model was able to compare

local and global implementations of the cross-homeostatic rule, and a primary contribution of my work is to show that a local variation on this family of plasticity rules also succeeds, but only when counterbalanced by standard homeostatic forces.

Finally, given emerging data that suggests that both timing and working memory rely on the internal dynamics of local cortical circuits, I use behavioral experiments to test a link between timing and working memory. Specifically, chapter 4 tests the hypothesis that prospective timing is computationally linked with working memory by performing two complementary behavioral experiments in humans. In the first task, participants performed a simple variant of the delayed match-to-sample task in which cue identity could be used to predict the timing of the appearance of the probe stimulus (and thus when the memory judgement would need to be made). In the second task, participants had to make a timing judgement of the delay duration to respond correctly, but remembering the cue identity was not required to perform the task. I found that participants learned the task-irrelevant timing information in the working memory task, and they stored task-irrelevant cue identity information in working memory in the explicit timing task. My results are consistent with the hypothesis that in some cases working memory and timing information are multiplexed in a time-varying format because of the importance of predicting when working memory will be used.

1.7 Chapter 1 References

1. Buonomano, D.V., *The biology of time across different scales*. 2007. p. 594-597.

2. Tallal, P. *In the Perception of Speech Time is of the Essence*. Berlin, Heidelberg: Springer Berlin Heidelberg.

3. Aasland, W.A. and S.R. Baum, *Temporal parameters as cues to phrasal boundaries: A comparison of processing by left- and right-hemisphere brain-damaged individuals.* Brain and Language, 2003. **87**(3): p. 385-399.

4. Shannon, R.V., et al., *Speech Recognition with Primarily Temporal Cues.* Science, 1995. **270**(5234): p. 303-304.

5. Drullman, R., *Temporal envelope and fine structure cues for speech intelligibility.* The Journal of the Acoustical Society of America, 1995. **97**(1): p. 585-592.

6. Mauk, M.D. and D.V. Buonomano, *The neural basis of temporal processing.* Annual review of neuroscience, 2004. **27**(1): p. 307-340.

7. Meck, W.H. and R.B. Ivry, *Editorial overview: Time in perception and action.* Current Opinion in Behavioral Sciences, 2016. **8**: p. 6-10.

8. Buzsáki, G. and R. Llinás, *Space and time in the brain.* Science, 2017. **358**(6362): p. 482-485.

9. Alder, T.B. and G.J. Rose, *Long-term temporal integration in the nauran auditory system.* Nature Neuroscience, 1998. **1**(6): p. 519-523.

10. Hebets, E.A. and D.R. Papaj, *Complex signal function: Developing a framework of testable hypotheses.* Behavioral Ecology and Sociobiology, 2005. **57**(3): p. 197-214.

11. Pollack, G., *Who, what, where? Recognition and localization of acoustic signals by insects.* Current Opinion in Neurobiology, 2000. **10**(6): p. 763-767.

12. Gerhardt, H.C. and F. Huber, *Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions*. 2002.

13. Thorson, J., T. Weber, and F. Huber, *Auditory Behavior of the Cricket - II. Simplicity of Calling-Song Recognition in Gryllus, and Anomalous Phonotaxis at Abnormal Carrier Frequencies.* Journal of Comparative Physiology A, 1982. **146**: p. 361-378.

14. Kostarakos, K. and B. Hedwig, *Calling Song Recognition in Female Crickets: Temporal Tuning of Identified Brain Neurons Matches Behavior.* Journal of Neuroscience, 2012. **32**(28): p. 9601-9612.

15. Carlson, B.A., *Electric signaling behavior and the mechanisms of electric organ discharge production in mormyrid fish.* Journal of Physiology Paris, 2002. **96**(5-6): p. 405-419.

16. Carlson, B.A. and C.D. Hopkins, *Central control of electric signaling behavior in the mormyrid Brienomyrus brachyistius: Segregation of behavior-specific inputs and the role of modifiable recurrent inhibition.* Journal of Experimental Biology, 2004. **207**(7): p. 1073-1084.

17. Covey, E. and J.H. Casseday, *Timing in the auditory system of the bat.* Annual Review of Physiology, 1999. **61**: p. 457-476.

18. Pai, S., et al., *Minimal Impairment in a Rat Model of Duration Discrimination Following Excitotoxic Lesions of Primary Auditory and Prefrontal Cortices.* Frontiers in Systems Neuroscience, 2011. **5**(September): p. 1-13.

19. Gouvêa, T.S., et al., *Striatal dynamics explain duration judgments.* eLife, 2015. **4**(December2015): p. 1-14.

20. Wright, B.A., et al., *Learning and generalization of auditory temporal-interval discrimination in humans.* Journal of Neuroscience, 1997. **17**(10): p. 3956-3963.

21. Bratzke, D., T. Seifried, and R. Ulrich, *Perceptual learning in temporal discrimination: Asymmetric cross-modal transfer from audition to vision.* Experimental Brain Research, 2012. **221**(2): p. 205-210.

22. Sakai, M., et al., *Neural mechanisms of interstimulus interval-dependent responses in the primary auditory cortex of awake cats.* BMC Neuroscience, 2009. **10**.

23. Goel, A. and D.V. Buonomano, *Timing as an intrinsic property of neural networks: evidence from in vivo and in vitro experiments.* Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 2014. **369**(1637): p. 20120460-20120460.

24. Rose, G.J. and R.R. Capranica, *Temporal selectivity in the central auditory system of the leopard frog.* Science, 1983. **219**(4588): p. 1087-1089.

25. Fortune, E.S. and G.J. Rose, *Passive and active membrane properties contribute to the temporal filtering properties of midbrain neurons in vivo.* Journal of Neuroscience, 1997. **17**(10): p. 3815-3825.

26. Pluta, S.R. and M. Kawasaki, *Temporal selectivity in midbrain electrosensory neurons identified by modal variation in active sensing.* Journal of Neurophysiology, 2010. **104**(1): p. 498-507.

27. Carlson, B.A., *Temporal-Pattern Recognition by Single Neurons in a Sensory Pathway Devoted to Social Communication Behavior.* Journal of Neuroscience, 2009. **29**(30): p. 9417-9428.

28. Baker, C.A. and B.A. Carlson, *Short-Term Depression, Temporal Summation, and Onset Inhibition Shape Interval Tuning in Midbrain Neurons.* Journal of Neuroscience, 2014. **34**(43): p. 14272-14287.

29. Baker, C.A., et al., *Behavioral and Single-Neuron Sensitivity to Millisecond Variations in Temporally Patterned Communication Signals.* Journal of Neuroscience, 2016. **36**(34): p. 8985-9000.

30. Casseday, J.H., D. Ehrlich, and E. Covey, *Neural tuning for sound duration: Role of inhibitory mechanisms in the inferior colliculus.* Science, 1994. **264**(5160): p. 847-850.

31. Fuzessery, Z.M. and J.C. Hall, *Sound duration selectivity in the pallid bat inferior colliculus.* Hearing Research, 1999. **137**(1-2): p. 137-154.

32. Aubie, B., R. Sayegh, and P.A. Faure, *Duration Tuning across Vertebrates.* Journal of Neuroscience, 2012. **32**(18): p. 6373-6390.

33. Pinheiro, A.D., M. Wu, and P.H.S. Jen, *Encoding repetition rate and duration in the inferior colliculus of the big brown bat, Eptesicus fuscus.* Journal of Comparative Physiology A, 1991. **169**(1): p. 69-85.

34. Faure, P.A., et al., *Temporal masking reveals properties of sound-evoked inhibition in duration-tuned neurons of the inferior colliculus.* Journal of Neuroscience, 2003. **23**(7): p. 3052-3065.

35. Potter, H.D., *PATTERNS OF ACOUSTICALLY EVOKED DISCHARGES OF NEURONS IN THE MESENCEPHALON OF THE BULLFROG.* Journal of Neurophysiology, 1965. **28**(6): p. 1155-1184.

36. Narins, P.M. and R.R. Capranica, *Neural Adaptations for Processing the Two-Note Call of the Puerto Rican Treefrog, Eleutherodactylus coqui.* Brain, Behavior and Evolution, 1980. **17**(1): p. 48-66.

37. Gooler, D.M. and A.S. Feng, *Temporal coding in the frog auditory midbrain: The influence of duration and rise-fall time on the processing of complex amplitude-modulated stimuli.* Journal of Neurophysiology, 1992. **67**(1): p. 1-22.

38. Brand, A., A. Urban, and B. Grothe, *Duration tuning in the mouse auditory midbrain.* Journal of Neurophysiology, 2000. **84**(4): p. 1790-1799.

39. Pérez-González, D., et al., *Duration selective neurons in the inferior colliculus of the rat: Topographic distribution and relation of duration sensitivity to other response properties.* Journal of Neurophysiology, 2006. **95**(2): p. 823-836.

40. Chen, G.D., *Effects of stimulus duration on responses of neurons in the chinchilla inferior colliculus.* Hearing Research, 1998. **122**(1-2): p. 142-150.

41. He, J., et al., *Temporal integration and duration tuning in the dorsal zone of cat auditory cortex.* Journal of Neuroscience, 1997. **17**(7): p. 2615-2625.

42. Duysens, J., S.J. Schaafsma, and G.A. Orban, *Cortical Off Response Tuning for Stimulus Duration.* Vision Research, 1996.

43. Doupe, A.J., *Song- and order-selective neurons in the songbird anterior forebrain emerge during vocal development.* Biomedical Research, 1997. **18**(SUPPL. 1): p. 17-37.

44. Zhou, X., et al., *Successive-signal biasing for a learned sound sequence.* Proceedings of the National Academy of Sciences, 2010. **107**(33): p. 14839-14844.

45. Konishi, M., *Effects of deafening on song development in American robins and black-headed grosbeaks.* Zeitschrift fur Tierpsychologie, 1965. **22**(5): p. 584-599.

46. Margoliash, D. and E.S. Fortune, *Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc.* The Journal of neuroscience : the official journal of the Society for Neuroscience, 1992. **12**(11): p. 4309-26.

47. Doupe, A.J. and P.K. Kuhl, *BIRDSONG AND HUMAN SPEECH: Common Themes and Mechanisms.* Annual Review of Neuroscience, 1999. **22**(1): p. 567-631.

48. Boari, S. and A. Amador, *Neural coding of sound envelope structure in songbirds.* Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 2018. **204**(3): p. 285-294.

49. Kilgard, M.P. and M.M. Merzenich, *Order-sensitive plasticity in adult primary auditory cortex.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(5): p. 3205-9.

50. Kilgard, M.P. and M.M. Merzenich, *Plasticity of temporal information processing in the primary auditory cortex.* Nature Neuroscience, 1998. **1**(8): p. 727-731.

51. Yin, P., et al., *Early stages of melody processing: Stimulus-sequence and task-dependent neuronal activity in monkey auditory cortical fields A1 and R.* Journal of Neurophysiology, 2008. **100**(6): p. 3009-3029.

52. Fishman, Y.I., et al., *Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey.* Hearing Research, 2001. **151**(1): p. 167-187.

53. Fishman, Y.I., J.C. Arezzo, and M. Steinschneider, *Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration.* J Acoust Soc Am, 2004. **116**(3): p. 1656-70.

54. Nelken, I., *Stimulus-specific adaptation and deviance detection in the auditory system: experiments and models.* Biol Cybern, 2014. **108**(5): p. 655-63.

55. Angeloni, C. and M.N. Geffen, *Contextual modulation of sound processing in the auditory cortex.* Current Opinion in Neurobiology, 2018. **49**: p. 8-15.

56. Das, A. and C.D. Gilbert, *Topography of contextual modulations mediated by short-range interactions in primary visual cortex.* Nature, 1999. **399**(6737): p. 655-661.

57. Asari, H. and A.M. Zador, *Long-lasting context dependence constrains neural encoding models in rodent auditory cortex.* Journal of Neurophysiology, 2009. **102**(5): p. 2638-2656.

58. Cohen-Kashi Malina, K., et al., *Imbalance between excitation and inhibition in the somatosensory cortex produces postadaptation facilitation.* Journal of Neuroscience, 2013. **33**(19): p. 8463-8471.

59. Phillips, D.P., S.E. Hall, and J.L. Hollett, *Repetition rate and signal level effects on neuronal responses to brief tone pulses in cat auditory cortex.* The Journal of the Acoustical Society of America, 1989. **85**(6): p. 2537-2549.

60. Wilson, D.A., *Odor specificity of habituation in the rat anterior piriform cortex.* Journal of Neurophysiology, 2000. **83**(1): p. 139-145.

61. Ahissar, E., et al., *Temporal frequency of whisker movement. II. Laminar organization of cortical representations.* J Neurophysiol, 2001. **86**(1): p. 354-67.

62. Chung, S., X. Li, and S.B. Nelson, *Short-term depression at thalamocortical synapses contributes to rapid adaptation of cortical sensory responses in vivo.* Neuron, 2002. **34**(3): p. 437-446.

63. Katz, Y., J.E. Heiss, and I. Lampl, *Cross-Whisker Adaptation of Neurons in the Rat Barrel Cortex.* The Journal of Neuroscience, 2006. **26**(51): p. 13363-13372.

64. Kohn, A., *Visual adaptation: Physiology, mechanisms, and functional benefits.* Journal of Neurophysiology, 2007. **97**(5): p. 3155-3164.

65. Gutnisky, D.A. and V. Dragoi, *Adaptive coding of visual information in neural populations.* Nature, 2008. **452**(7184): p. 220-224.

66. Natan, R.G., et al., *Complementary control of sensory adaptation by two types of cortical interneurons.* eLife, 2015. **4**(OCTOBER2015): p. 1-27.

67. Calford, M.B. and M.N. Semple, *Monaural inhibition in cat auditory cortex.* Journal of Neurophysiology, 1995. **73**(5): p. 1876-1891.

68. Brosch, M. and C.E. Schreiner, *Time course of forward masking tuning curves in cat primary auditory cortex.* J Neurophysiol, 1997. **77**(2): p. 923-43.

69. Wehr, M. and A.M. Zador, *Synaptic mechanisms of forward suppression in rat auditory cortex.* Neuron, 2005. **47**(3): p. 437-445.

70. Phillips, E.A.K., C.E. Schreiner, and A.R. Hasenstaub, *Diverse effects of stimulus history in waking mouse auditory cortex.* Journal of Neurophysiology, 2017. **118**(2): p. 1376-1393.

71. Phillips, E.A.K., C.E. Schreiner, and A.R. Hasenstaub, *Cortical Interneurons Differentially Regulate the Effects of Acoustic Context.* Cell Reports, 2017. **20**(4): p. 771-778.

72. Brosch, M., A. Schulz, and H. Scheich, *Processing of sound sequences in macaque auditory cortex: Response enhancement.* Journal of Neurophysiology, 1999. **82**(3): p. 1542-1559.

73. Brosch, M. and C.E. Schreiner, *Sequence sensitivity of neurons in cat primary auditory cortex.* Cerebral cortex (New York, N.Y.: 1991), 2000. **10**(12): p. 1155-1167.

74. Bartlett, E.L. and X. Wang, *Long-lasting modulation by stimulus context in primate auditory cortex.* Journal of Neurophysiology, 2005. **94**(1): p. 83-104.

75. Sadagopan, S. and X. Wang, *Nonlinear spectrotemporal interactions underlying selectivity for complex sounds in auditory cortex.* Journal of Neuroscience, 2009. **29**(36): p. 11192-11202.

76. Natan, R.G., W. Rao, and M.N. Geffen, *Cortical Interneurons Differentially Shape Frequency Tuning following Adaptation.* Cell Reports, 2017. **21**(4): p. 878-890.

77. Fiala, J.C., S. Grossberg, and D. Bullock, *Metabotropic Glutamate Receptor Activation in Cerebellar Purkinje Cells as Substrate for Adaptive Timing of the Classically Conditioned Eye-Blink Response.* The Journal of Neuroscience, 1996. **16**(11): p. 3760-3774.

78. Hooper, S.L., E. Buchman, and K.H. Hobbs, *A computational role for slow conductances: single-neuron models that measure duration.* Nature Neuroscience, 2002. **5**(6): p. 552-556.

79. Kohashi, T. and B.A. Carlson, *A fast BK-type KCa current acts as a postsynaptic modulator of temporal selectivity for communication signals.* Frontiers in Cellular Neuroscience, 2014. **8**(September): p. 1-16.

80. Buonomano, D.V. and M.M. Merzenich, *Temporal information transformed into a spatial code by a neural network with realistic properties.* Science (New York, N.Y.), 1995. **267**(5200): p. 1028-1030.

81. Buonomano, D.V., *Decoding temporal information: A model based on short-term synaptic plasticity.* The Journal of neuroscience : the official journal of the Society for Neuroscience, 2000. **20**(3): p. 1129-41.

82. Fortune, E.S. and G.J. Rose, *Short-term synaptic plasticity as a temporal filter.* Trends in Neurosciences, 2001. **24**(7): p. 381-385.

83. George, A.A., et al., *A Diversity of Synaptic Filters Are Created by Temporal Summation of Excitation and Inhibition.* Journal of Neuroscience, 2011. **31**(41): p. 14721-14734.

84. Rose, G.J., C.J. Leary, and C.J. Edwards, *Interval-counting neurons in the anuran auditory midbrain: Factors underlying diversity of interval tuning.* Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 2011. **197**(1): p. 97-108.

85. Alluri, R.K., et al., *Phasic, suprathreshold excitation and sustained inhibition underlie neuronal selectivity for short-duration sounds.* Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(13): p. E1927-E1935.

86. Aubie, B., S. Becker, and P.A. Faure, *Computational models of millisecond level duration tuning in neural circuits.* Journal of Neuroscience, 2009. **29**(29): p. 9255-9270.

87. Large, E.W. and J.D. Crawford, *Auditory temporal computation: Interval selectivity based on post-inhibitory rebound.* Journal of Computational Neuroscience, 2002. **13**(2): p. 125-142.

88. Zucker, R.S. and W.G. Regehr, *Short-Term Synaptic Plasticity.* Annual Review of Physiology, 2002. **64**(1): p. 355-405.

89. Markram, H., Y. Wang, and M. Tsodyks, *Differential signaling via the same axon of neocortical pyramidal neurons.* Proc Natl Acad Sci U S A, 1998. **95**(9): p. 5323-8.

90. Jackman, S.L., et al., *The calcium sensor synaptotagmin 7 is required for synaptic facilitation.* Nature, 2016. **529**(7584): p. 88-91.

91. Beierlein, M., J.R. Gibson, and B.W. Connors, *Two Dynamically Distinct Inhibitory Networks in Layer 4 of the Neocortex.* Journal of Neurophysiology, 2003. **90**(5): p. 2987-3000.

92. Reyes, A., et al., *Target-cell-specific facilitation and depression in neocortical circuits.* Nature neuroscience, 1998. **1**(4): p. 279-285.

93. Gupta, A., *Organizing Principles for a Diversity of GABAergic Interneurons and Synapses in the Neocortex.* Science, 2000. **287**(5451): p. 273-278.

94. Ma, Y., H. Hu, and A. Agmon, *Short-Term Plasticity of Unitary Inhibitory-to-Inhibitory Synapses Depends on the Presynaptic Interneuron Subtype.* Journal of Neuroscience, 2012. **32**(3): p. 983-988.

95. Viaene, A.N., I. Petrof, and S.M. Sherman, *Synaptic Properties of Thalamic Input to the Subgranular Layers of Primary Somatosensory and Auditory Cortices in the Mouse.* Journal of Neuroscience, 2011. **31**(36): p. 12738-12747.

96. Reyes, A.D., *Synaptic short-term plasticity in auditory cortical circuits.* Hearing Research, 2011. **279**(1-2): p. 60-66.

97. Atzori, M., et al., *Differential synaptic processing separates stationary from transient inputs to the auditory cortex.* Nature Neuroscience, 2001. **4**(12): p. 1230-1237.

98. Reyes, a. and B. Sakmann, *Developmental switch in the short-term modification of unitary EPSPs evoked in layer 2/3 and layer 5 pyramidal neurons of rat neocortex.* The Journal of neuroscience : the official journal of the Society for Neuroscience, 1999. **19**(10): p. 3827-3835.

99. Kumar, S.S.S. and J.R.R. Huguenard, *Properties of excitatory synaptic connections mediated by the corpus callosum in the developing rat neocortex.* Journal of Neurophysiology, 2001. **86**(6): p. 2973-2985.

100. Oswald, A.M.M.M. and A.D. Reyes, *Maturation of Intrinsic and Synaptic Properties of Layer 2/3 Pyramidal Neurons in Mouse Auditory Cortex.* Journal of Neurophysiology, 2008. **99**(6): p. 2998-3008.

101. Finnerty, G.T., L.S.E. Roberts, and B.W. Connors, *Sensory experience modifies the short-term dynamics of neocortical synapses.* Nature, 1999. **400**(6742): p. 367-371.

102. Diaz-Quesada, M., et al., *Diverse Thalamocortical Short-Term Plasticity Elicited by Ongoing Stimulation.* Journal of Neuroscience, 2014. **34**(2): p. 515-526.

103. Hsieh, C.Y., S.J. Cruikshank, and R. Metherate, *Differential modulation of auditory thalamocortical and intracortical synaptic transmission by cholinergic agonist.* Brain Research, 2000. **880**(1-2): p. 51-64.

104. Campagnola, L., et al., *Connectivity and Synaptic Physiology in the Mouse and Human Neocortex.* bioRxiv, 2021.

105. Reyes, A.D., *Synaptic short-term plasticity in auditory cortical circuits.* Hearing Research, 2011. **279**(1–2): p. 60-66.

106. Levy, R.B. and A.D. Reyes, *Spatial Profile of Excitatory and Inhibitory Synaptic Connectivity in Mouse Primary Auditory Cortex.* Journal of Neuroscience, 2012. **32**(16): p. 5609-5619.

107. Takesian, A.E., et al., *Hearing loss differentially affects thalamic drive to two cortical interneuron subtypes.* Journal of Neurophysiology, 2013. **110**(4): p. 999-1008.

108. Tan, Z., et al., *Robust but delayed thalamocortical activation of dendritic-targeting inhibitory interneurons.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(6): p. 2187-2192.

109. Abbott, L.F. and W.G. Regehr, *Synaptic computation.* Nature, 2004. **431**(7010): p. 796-803.

110. Anwar, H., et al., *Functional roles of short-term synaptic plasticity with an emphasis on inhibition.* Current Opinion in Neurobiology, 2017. **43**: p. 71-78.

111. Abbott, L.F., et al., *Synaptic depression and cortical gain control.* Science (New York, N.Y.), 1997. **275**(5297): p. 220-224.

112. Cook, D., et al., *Synaptic depression in the localization of sound.* Nature, 2003. **421**: p. 66-70.

113. Rotman, Z., P.Y.Y. Deng, and V.A. Klyachko, *Short-Term Plasticity Optimizes Synaptic Information Transmission.* Journal of Neuroscience, 2011. **31**(41): p. 14800-14809.

114. Rosenbaum, R., J. Rubin, and B. Doiron, *Short term synaptic depression imposes a frequency dependent filter on synaptic information transfer.* PLoS Computational Biology, 2012. **8**(6).

115. Markram, H., Y. Wang, and M. Tsodyks, *Differential signaling via the same axon of neocortical pyramidal neurons.* Proc. Natl. Acad. Sci. USA., 1998. **95**: p. 5323-5328.

116. Buonomano, D.V., *Decoding temporal information: a model based on short-term synaptic plasticity.* J Neurosci., 2000. **20**: p. 1129-1141.

117. Rose, G.J., *Time computations in anuran auditory systems.* Frontiers in Physiology, 2014. **5 MAY**(May): p. 1-7.

118. Edwards, C.J., T.B. Alder, and G.J. Rose, *Auditory midbrain neurons that count.* Nature Neuroscience, 2002. **5**(10): p. 934-936.

119. Edwards, C.J., C.J. Leary, and G.J. Rose, *Counting on Inhibition and Rate-Dependent Excitation in the Auditory System.* Journal of Neuroscience, 2007. **27**(49): p. 13384-13392.

120. Edwards, C.J., C.J. Leary, and G.J. Rose, *Mechanisms of Long-Interval Selectivity in Midbrain Auditory Neurons: Roles of Excitation, Inhibition, and Plasticity.* Journal of Neurophysiology, 2008. **100**(6): p. 3407-3416.

121. Park, Y. and M.N. Geffen, *A circuit model of auditory cortex.* PLOS Computational Biology, 2020. **16**(7): p. e1008016.

122. Seay, M.J., et al., *Differential Short-Term Plasticity of PV and SST Neurons Accounts for Adaptation and Facilitation of Cortical Neurons to Auditory Tones.* The Journal of neuroscience : the official journal of the Society for Neuroscience, 2020. **40**(48): p. 9224-9235.

123. Maass, W., T. Natschläger, and H. Markram, *Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations.* Neural Computation, 2002. **14**(11): p. 2531-2560.

124. Buonomano, D.V. and W. Maass, *State-dependent computations: spatiotemporal processing in cortical networks.* Nature reviews. Neuroscience, 2009. **10**(2): p. 113-125.

125. Nikolić, D., et al., *Distributed Fading Memory for Stimulus Properties in the Primary Visual Cortex.* PLoS Biology, 2009. **7**(12): p. e1000260-e1000260.

126. Karmarkar, U.R. and D.V. Buonomano, *Timing in the Absence of Clocks: Encoding Time in Neural Network States.* Neuron, 2007. **53**(3): p. 427-438.

127. Spencer, R.M.C., U. Karmarkar, and R.B. Ivry, *Evaluating dedicated and intrinsic models of temporal encoding by varying context.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2009. **364**(1525): p. 1853-1863.

128. Pérez, O. and H. Merchant, *The synaptic properties of cells define the hallmarks of interval timing in a recurrent neural network.* The Journal of Neuroscience, 2018: p. 2651-17.

129. P. Carvalho, T., et al., *A novel learning rule for long-term plasticity of short-term synaptic plasticity enhances temporal processing.* Frontiers in Integrative Neuroscience, 2011. **5**(May): p. 20-20.

130. Lee, T.P. and D.V. Buonomano, *Unsupervised formation of vocalization-sensitive neurons: A cortical model based on short-term and homeostatic plasticity.* Neural Computation, 2012. **24**(10): p. 2579-2603.

Chapter 2: Differential Short-Term Plasticity of PV and SST Neurons Accounts for Adaptation and Facilitation of Cortical Neurons to Auditory Tones

Published as: Seay MJ, Natan RG, Geffen MN, Buonomano DV. *Differential Short-Term Plasticity of PV and SST Neurons Accounts for Adaptation and Facilitation of Cortical Neurons to Auditory Tones*, 2020. Journal of Neuroscience. 40(48): p. 9224-9235.

Abstract

Cortical responses to sensory stimuli are strongly modulated by temporal context. One of the best studied examples of such modulation is sensory adaptation. We first show that in response to repeated tones pyramidal (Pyr) neurons in male mouse primary auditory cortex (A1) exhibit facilitating and stable responses, in addition to adapting responses. To examine the potential mechanisms underlying these distinct temporal profiles, we developed a reduced spiking model of sensory cortical circuits that incorporated the signature short-term synaptic plasticity (STP) profiles of the inhibitory Parvalbumin (PV) and Somatostatin (SST) interneurons. The model accounted for all three temporal response profiles as the result of dynamic changes in excitatory/inhibitory balance produced by STP, primarily through shifts in the relative latency of Pyr and inhibitory neurons. Transition between the three response profiles was possible by changing the strength of the inhibitory PV→Pyr and SST→Pyr synapses. The model predicted that a unit's latency would be related to its temporal

profile. Consistent with this prediction, the latency of stable units was significantly shorter than that of adapting and facilitating units. Furthermore, because of the history-dependence of STP the model generated a paradoxical prediction: that inactivation of inhibitory neurons during one tone would decrease the response of A1 neurons to a subsequent tone. Indeed, we observed that optogenetic inactivation of PV neurons during one tone counterintuitively decreased the spiking of Pyr neurons to a subsequent tone 400 ms later. These results provide evidence that STP is critical to temporal context-dependent responses in the sensory cortex.

Chapter 2: Differential Short-Term Plasticity of PV and SST Neurons Accounts for Adaptation and Facilitation of Cortical Neurons to Auditory Tones

2.1 Introduction

Auditory processing requires identification of temporal structure of stimuli on the subsecond time scale, including order, duration, interval, and temporal context. Speech, for example, is characterized not only by its spectral structure, but its temporal structure—including the order and duration of phonemes as well as intervals between them [1-3]. Despite the importance of identifying the temporal structure of auditory stimuli, the neural correlates and mechanisms of this process remain poorly understood.

One of the best studied examples of the modulation of the responses of auditory neurons by temporal context is sensory adaptation, in which responses to pairs or trains of consecutive tones progressively decrease [4-13]. Sensory adaptation is observed in the somatosensory and visual modalities as well [14-20]. There is, however, significant diversity in the modulation of cortical responses by consecutive stimuli, including neurons that exhibit enhanced responses to specific spatiotemporal patterns [4, 5, 11, 13, 21, 22]. The diversity of adapting and facilitating responses presumably reflects the cross-purpose need to habituate to repetitive stimuli that carry little information (e.g., the tic of a clock), *and* to detect critical information carried by events occurring late in a sequence (e.g. in Morse code the letter *I* is represented by 2 consecutive dots, and the letter *H* by 4 dots) [23-25].

Rodent studies have focused primarily on adapting (i.e. decreasing) responses to sequences of tones and proposed a number of potential underlying mechanisms

including short-term synaptic depression and long-lasting inhibition [7, 12, 23]. More recent studies have suggested differential roles for Parvalbumin (PV) and Somatostatin (SST) inhibitory interneurons in sensory adaptation [6, 26] and forward suppression measured with tone pairs [22, 27]. Auditory cortical neurons, however, exhibit a range of temporal profiles in response to sequences of tones, including stable responses and progressive facilitation. The mechanisms responsible for this diversity of temporal profiles are not understood, and it remains an open question whether the adapting, stable, and facilitating responses can be explained by the same set of mechanisms or rely on fundamentally different properties.

Previous research has suggested that short-term synaptic plasticity (STP) plays an important role in governing sensitivity to temporal context [16, 28-32]. Here we characterized the different temporal profiles of auditory neurons to sequences of tones, and developed a spiking neuron model that incorporates known differential short-term synaptic dynamics of PV and SST interneurons. In contrast to previous firing rate models [26, 27], here we use a spike-based model that accurately captures STP dynamics and the relative timing between excitation and inhibition, and makes experimental predictions about spike latency. We found that both adaptation and facilitation can be explained in terms of differences in the relative balance of inhibition originating from PV and SST neurons. The model generated a number of predictions including: 1) Different temporal profiles should be correlated with specific firing latency signatures, and 2) inactivating inhibitory neurons should prevent normally occurring STP of IPSPs, and thus alter pyramidal (Pyr) neuron responses to subsequent tones. Both predictions were tested and supported by the experimental evidence.

2.2 Results

2.2.1 A1 neurons exhibit distinct temporal profiles in response to repeated tones.

We first examined the temporal response profiles of primary auditory cortex (A1) neurons to trains of eight consecutive tones presented at a 2.5 Hz (400 ms). From a total of 406 neurons across 11 animals (5 PV-Cre and 6 SST-Cre), we identified a total of 1486 neuron-frequency pairs that exhibited a significant evoked response to either the first or last tone (see Methods). Based on the temporal profile of the number of spikes elicited in response to each of the eight tones in each train, we classified each neuron-frequency pair as adapting, stable, or facilitating (see Methods). Of these 38.7% were adapting, 55.5% stable, and 5.8% facilitating. **Fig. 2.1** displays sample raster plots of each type of temporal profile (**Fig. 2.1A, C, and E**) and the population post-stimulus time histograms (PSTHs) across all neuron-frequency pairs within each category (**Fig. 2.1B, D, and F**). Consistent with previous studies these results confirm that while many neurons exhibit robust sensory adaptation, there is significant diversity in the temporal profile of the responses including neurons that exhibit robust facilitation.

2.2.2 Model of a simple cortical microcircuit that incorporates short-term synaptic plasticity

The presence of adaptation or facilitation within cortical neurons establishes that responses are sensitive to temporal context—in other words that there is a memory of recent stimulus history in the circuits. Early models suggested that one of the mechanisms underlying the sensitivity to temporal context is STP [16, 28, 29, 31, 33-36], and recent work has focused on the differential STP among different classes of inhibitory neurons as contributing to sensory adaptation[6, 22]. We thus developed

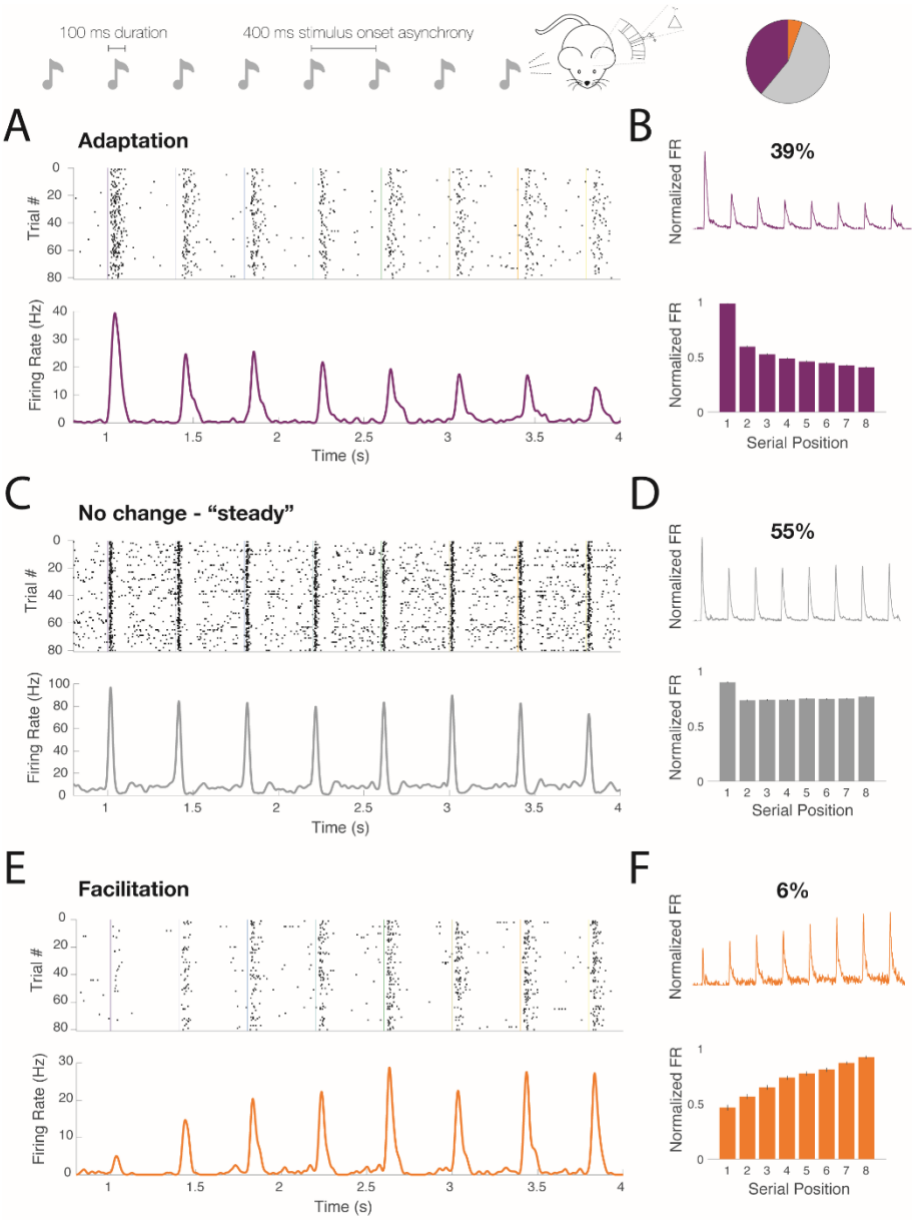

Figure 2.1 Single units in A1 exhibit diverse temporal profiles of responses to sequences of repeated tones, including adaptation, no change, and facilitation.
In the experiment, a train of eight consecutive repetitions of the same 100 ms pure tone stimulus was presented at a rate of 2.5 Hz. The pie chart at top right shows the proportions of neuron-frequency pairs that exhibited adaptation (purple), no change (grey), or facilitation (orange). **A**. Spike raster (upper) and post-stimulus time histogram (PSTH; lower) for a neuron-frequency pair that exhibits classical *adaptation*, in which sensory responses to the same physical stimulus *decrease* with repetition on short timescales. **B**. 39% of neuron-frequencies with significant evoked activity were adapting. Population average normalized PSTH (upper) and bar plot of average normalized firing rate within 10-70 ms following tone onset by serial position (lower). **C**. Spike raster (upper) and PSTH (lower) for a neuron-frequency pair that exhibits *no change* or a *steady* firing rate over repetition. **D**. 55% of neuron-frequencies with significant evoked activity exhibited no change in firing rate over repetition. **E**. Spike raster (upper) and PSTH (lower) for a neuron-frequency pair that exhibits *facilitation*, in which sensory responses to the same physical stimulus *increase* with repetition on short timescales. **F**. 6% of neuron-frequencies with significant evoked activity were facilitating.

a spike-based model of feedforward activation of pyramidal neurons that incorporated the experimentally characterized short-term synaptic plasticity of PV and SST interneurons (**Fig. 2.2**).

Importantly, we used published empirically-derived STP estimates for each of the five synapses in our Pyr-PV-SST microcircuit (**Fig. 2.2A**). We modeled mild depression at the Input→Pyr synapse [16, 37, 38], moderate depression at the Input→PV synapse [38-40], strong facilitation at the Input→SST synapse [39, 41], strong depression at the PV→Pyr synapse [40, 42, 43], and stable synaptic strength at the SST→Pyr synapse [44, 45]. STP parameters are shown in **Table 2.3**. Additionally, the intrinsic properties of the three neuron classes were based on aggregate estimates from the NeuroElectro project and select electrophysiological studies [46, 47]. Using these units and synapses, we assembled a circuit with dual disynaptic feedforward inhibition from both PV and SST onto the Pyr unit and simulated the experimental protocol (**Fig. 2.1**).

2.2.3 Model can account for all three temporal profiles by changing the inhibitory weights

We next asked whether this simple model with empirically-based STP values at the five synapses could account for all three experimentally observed temporal profiles. Our goal was to determine if the diversity of temporal profiles could be reproduced without changing the temporal dynamics of STP at each of the five synapses modeled here. Thus the free parameters in the model were the five synaptic weights: Input→Pyr, Input→PV, Input→SST, PV→Pyr, and SST→Pyr. We omitted higher order lateral connections, such as the SST→PV connections, because these are unlikely contribute to the fast latency responses studied here (see Methods). The input weights onto all

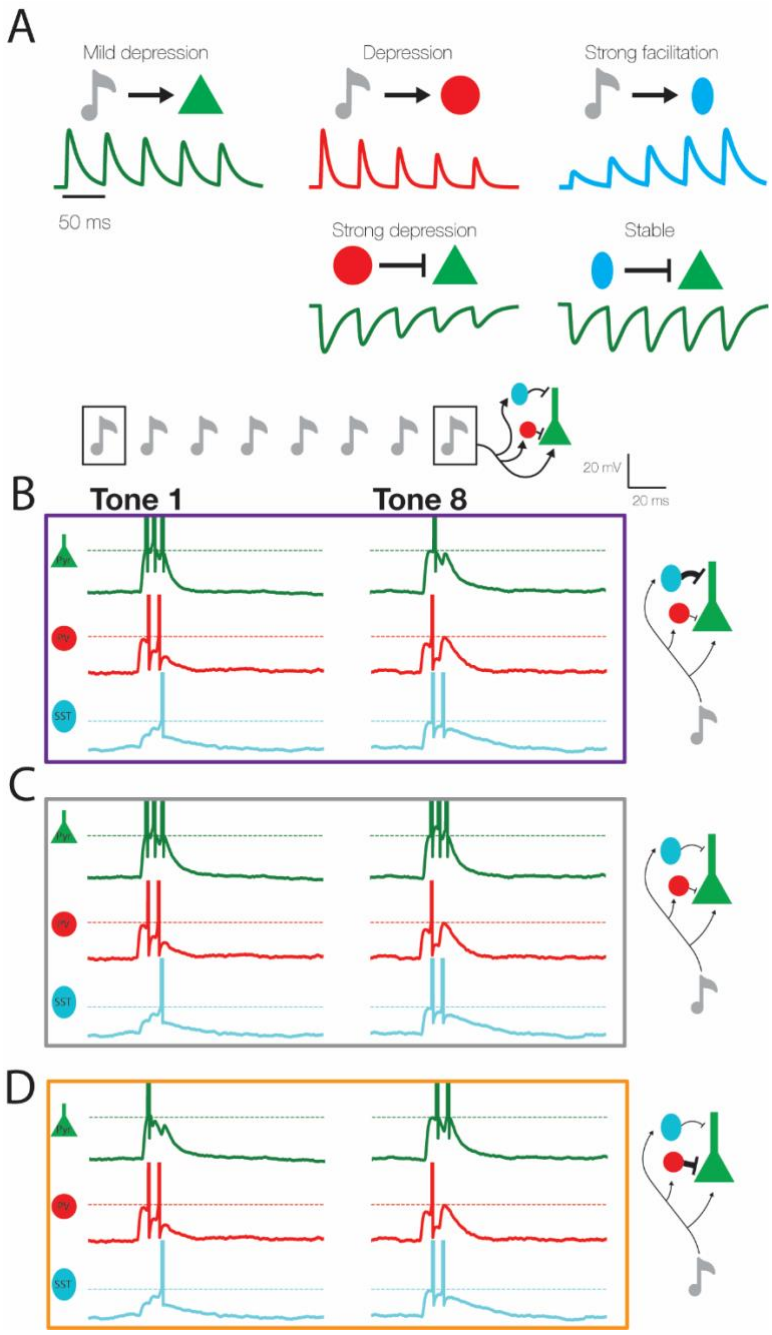

Figure 2.2 Spiking model of feedforward cortical microcircuit with empirically-based short-term synaptic plasticity (STP).
A. Three distinct units were modeled to resemble cortical pyramidal (Pyr, green), fast-spiking Parvalbumin-expressing interneurons (PV, red), and low-threshold-spiking Somatostatin-expressing interneurons (SST, cyan). The change in synaptic currents caused by repeated presynaptic spikes was governed by short-term synaptic plasticity (STP) derived from experimental observations. **B.** Single-unit membrane voltages from a model simulation of the experiment considered here (Natan et al., 2017). Because the SST→Pyr synapse is strong, Pyr unit spiking is suppressed during the 8th tone. **C.** Single-unit membrane voltages when both the PV→Pyr and SST→Pyr synapses are relatively weak and balanced. Pyr unit spiking is relatively unaffected. **D.** Single-unit membrane voltages when the PV→Pyr synapse is strong. Pyr unit spiking is strongly suppressed during the 1st tone but only weakly suppressed during the 8th tone, resulting in facilitation.

three neuron classes were constrained such that they elicited biologically reasonable firing rates in response to single tones in the absence of any inhibition. We thus anchored the input weights and focused on a parametric analysis of the weights of the PV→Pyr and SST→Pyr connections. As shown in **Fig. 2.2B-D,** it is possible to transition between all three temporal profiles by changing only the PV→Pyr and SST→Pyr weights. Starting in a regime with weak inhibitory weights from both PV and SST, we observe a steady response profile (**Fig. 2.2C**). This is intuitive because in the absence of strong inhibition the temporal profile is primarily shaped by STP of the Input→Pyr connection—which is weakly depressing. By increasing the strength of the SST→Pyr connection, the system shifts to the adapting profile (**Fig. 2.2B**). Although early in the train SST generated inhibition is relatively weak, increasing the strength of the PV→Pyr connection shifts the profile from adapting to facilitating (**Fig. 2.2D**).

Fig. 2.3 contrasts experimental (**Fig. 2.3A$_1$-C$_1$**) and simulated (**Fig. 2.3A$_2$-C$_2$**) examples of units from all three classes of temporal profiles. The unfilled bar plots display the model results of average evoked firing rates of the Pyr unit across twenty trials with independent noise currents provided to each unit. Importantly, the only parameters that were varied between the three simulated temporal profiles were the synaptic strengths (i.e. maximal conductances) of the SST→Pyr and the PV→Pyr synapses. All other parameters, including the STP parameters at each synapse, were the same across all simulations.

In order to quantify the robustness of the above results across different relative ratios of PV and SST inhibition we conducted a two-dimensional parameter search across the SST→Pyr and PV→Pyr weights. At each combination of inhibitory weights,

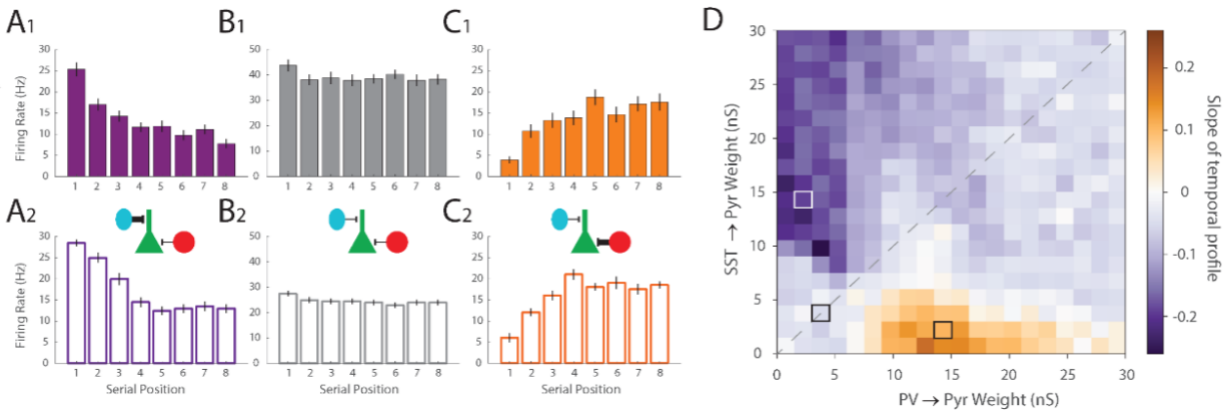

Figure 2.3 Model circuit reproduces experimentally observed adaptation, steady responses, and facilitation by changing relative strength of PV/SST inhibition.
These three temporal profiles reflect three of the weight sets across a systematic parametric analysis of the PV→Pyr and SST→Pyr weights (D). In the A, B, and C panels the upper filled bar plot displays the average tone-evoked firing rate from 10-70 ms for an exemplary experimentally-observed neuron-frequency pair across a minimum of 40 trials, while the bottom, unfilled bar plot displays the average tone-evoked firing rate of the Pyr unit in the model across 20 trials with independent noise when synaptic weights were set as indicated by the inset diagrams and the outlined weight combinations shown in D. All error lines indicate the standard error of the mean. **A1**. Exemple of an experimentally observed adapting response. **A2**. Simulated adapting Pyr unit. Bars indicate the mean firing rate evoked by a simulated sequence of "tones" based on their serial position. In this simulation, the SST→Pyr synapse was relatively strong, as indicated by the inset circuit diagram. **B1**. Experimental example of a steady neuron. **B2**. Simulated steady Pyr unit: both the PV→Pyr and SST→Pyr synapses were relatively weak and balanced, as indicated by the inset circuit diagram. **C1**. Experimental example of a facilitating Pyr neuron. **C2**. Simulated facilitating Pyr unit: the PV→Pyr synapse was relatively strong, as indicated by the inset circuit diagram. **D**. Color-coded heatmap of the slope of the Pyr firing rate across serial positions (i.e. the temporal profile) as the weights PV→Pyr and SST→Pyr synapses were parametrically varied. More intensely purple squares reflect adaptation, while more intense orange squares reflect facilitation.

we quantified the average firing rate of the Pyr unit at each serial position across 20 trials. To quantify and visualize the temporal profiles, we regressed the Pyr unit firing rate against the serial position and took the slope as a quantitative index in which positive values indicated facilitation, negative values adaptation, and values near zero indicated a steady firing profile (**Fig. 2.3D**). These results confirm the robustness of the model across different inhibitory weights, and highlight the importance of the relative balance of PV and SST inhibition. One can see that the diagonal of the parametric analysis reveals mostly steady responses, indicating that it is not simply the absolute strength of PV or SST inputs that determines the temporal profile. For example, at

PV→Pyr weights of 10-15 nS, facilitating, steady, or depressing profiles can be observed, depending on the weight of the SST→Pyr connection.

2.2.4 STP acts via changes in spike latency

The above results are driven by the dynamic shifts in E/I balance produced by STP. For example, facilitation of the Input→SST is responsible for the progressive decrease in Pyr firing in response to repeated tones. It is reasonable to assume that this modulation relies on an increase in the number of spikes in the SST neuron. A detailed analysis of the model, however, revealed a more complex mechanism. Specifically, much of the modulation of the firing rates of the Pyr neurons is not governed by the change in the spike number of inhibitory neurons, but by the shift in their firing latency.

To examine the importance of STP-dependent shifts in spike latency of inhibitory neurons in shaping Pyr profiles we first considered a circuit in the adaptation regime (**Fig. 2.4A**). As the Pyr unit firing rate decreases across serial positions, the 1st spike latency of the SST unit decreases in a correlated manner, due to short-term facilitation of the Input→SST synapse. Although the SST unit's 1st spike latency and the Pyr unit's firing rate simultaneously change in a nonlinear manner, there was a strong linear correlation between the two (r = 0.99, p < .001). In a circuit in the facilitation regime (**Fig. 2.4B**) the increase in Pyr unit firing rate across serial positions was correlated with an increase in 1st spike latency of the PV unit due to short-term depression of the Input→PV synapse (r = 0.96, p < .001)—note however, that the range of the latency shifts was narrower because of the relatively weak short-term depression at the Input→PV connection. Importantly, we note that such a relationship will only be found for circuits with a single dominant source of inhibition, and that in scenarios with weak or

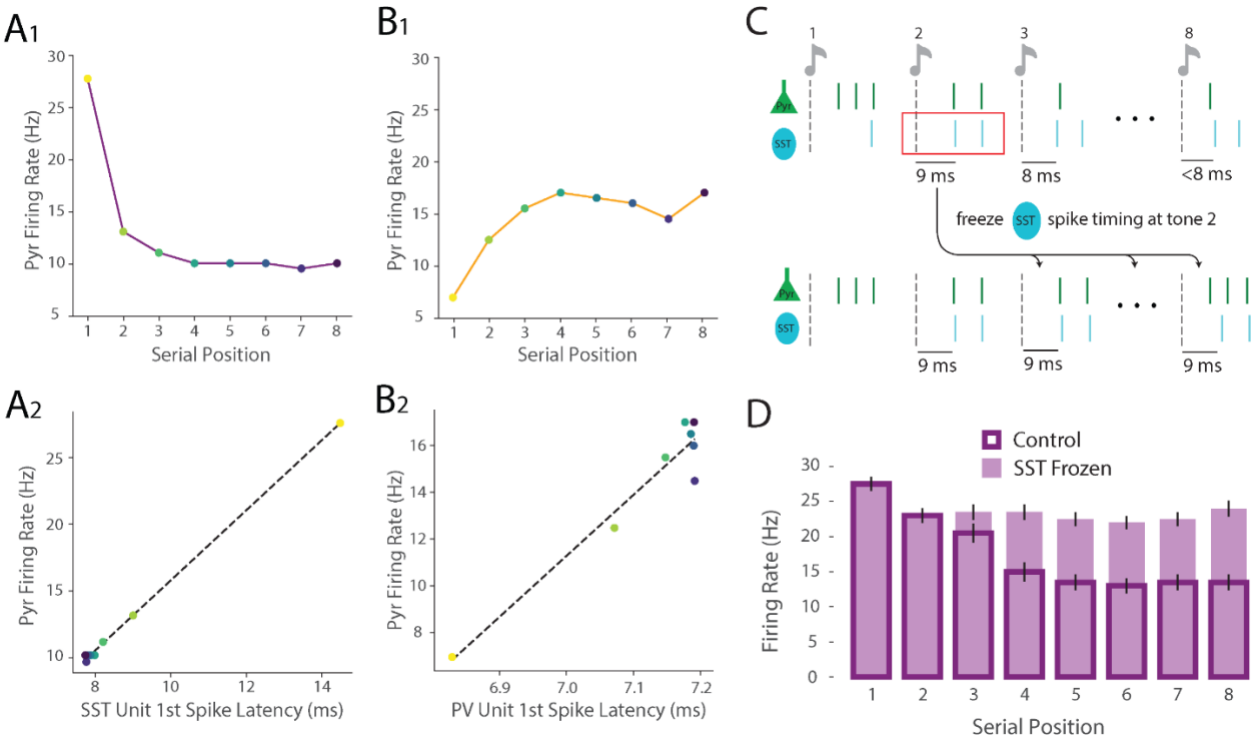

Figure 2.4 Temporal profiles are shaped by STP-driven changes in spike latency. **A1**. Decreasing firing rate in Pyr unit in a circuit in the adaptation regime across consecutive tones. **A2.** Relationship between firing rate in A1 and first spike latency of SST. Note that the color or each point corresponds to the serial position in A1. **B1**. Increasing firing rate in Pyr unit in a circuit in the facilitating regime across consecutive tones. Note the relatively narrow range of PV latency changes compared to SST. **B2**. Relationship between firing rate in A1 and first spike latency of PV. **C.** Starting from the model results in the adaptation weight regime, we recorded the average tone-evoked latency of SST spikes during the second tone. Then, we re-ran the simulation while artificially replacing the SST unit spiking on tones 3-8 to be the same as the average on tone 2. **D**. When SST unit spiking was "frozen" at its latency during tone 2, adaptation was eliminated despite no change in the firing rate of the SST unit.

balanced sources of inhibition (which tend to exhibit steady temporal profiles) such a relationship will not hold. The previous analyses suggest that decreases in the latency of SST spiking produced by short-term facilitation of the Input→SST synapse plays a role in the adaptation profile. But they do not demonstrate that the change in spike timing *causes* the change in Pyr unit firing rate within the model.

To demonstrate a causal relationship, we developed an approach for artificially altering the timing of the inhibitory unit. We reran the simulations under the adapting regime, but replaced the spike times of the SST unit for tones 3-8 with the average

spike times of the SST unit during tone 2 (**Fig. 2.4C**). Thus we artificially “froze” the SST unit spike times and prevented the STP of the Input→SST synapses from naturally decreasing the tone-evoked latency of the SST neurons, while preserving the increase in spike number produced by STP during tones. This manipulation caused the Pyr unit to exhibit a steady firing rate across serial positions 2-8, eliminating the adaptation after tone 2 (**Fig. 2.4D**). Importantly, the firing rate of the SST unit in the trials where its spike timing was “frozen” was identical to its firing rate shown in **Fig. 2.3A$_2$**. Thus, the temporal profile of adaptation can primarily be attributed to a progressive reduction in the tone-evoked spike latencies of the SST unit.

This result is consistent with previous experimental and computational data establishing that the temporal relationship between the onsets of excitatory postsynaptic potentials (EPSPs) and inhibitory postsynaptic potentials (IPSPs) in pyramidal neurons plays a fundamental role in shaping their neuronal responses [48, 49]. The earlier the onset of the IPSP, the more effective it is at preventing the EPSP from driving the pyramidal neuron to threshold. Our model predicts that the adaptation produced by progressive facilitation of the Input→SST synapse is not driven primarily by increased spiking of the SST neurons, but rather the progressive decrease in the latency of input-evoked SST spiking. It is relevant to stress that this latency effect would likely not be captured as well by firing rate models which are less well suited to pick up on the highly nonlinear interactions between shifts in EPSP/IPSP latencies and all-or-none spike generation [48-51].

2.2.5 Model correctly predicts longer response latencies under adaptation and facilitation

Relative timing of the EPSPs and IPSPs onto the pyramidal neurons is a critical determinant of the model. To be effective, inhibition must arrive quickly enough to interact with excitatory inputs and prevent potential spiking. Based on this observation, we hypothesized that steady neurons do not undergo strong temporal-context modulation, because they often fire before the onset of inhibition. Thus, a key prediction of our model is that there should be a relationship between the input-evoked latency of the Pyr unit and the extent to which that unit's firing rate is decreased by inhibition. More specifically, we predicted that short-latency Pyr units should be less vulnerable to inhibition while still exhibiting mild forms of firing rate adaptation due to mild depression at the Input→Pyr synapse. In contrast, we predicted that longer latency Pyr units should be more strongly modulated by feedforward inhibition.

To test these predictions, we first measured the response latency of each neuron-frequency pair in the experimental data. We operationalized response latency as the time of the maximal peak in the post-stimulus time histogram (PSTH) in response to the first (adapting and stable responses) or final (facilitating responses) tone. Even within a temporal profile class response latency varied considerably, as evidenced by a plot of PSTHs sorted by latency (**Fig. 2.5A-B**). However, consistent with our prediction, the mean latency of the steady group was significantly shorter than that of the adaptation and facilitation groups (**Fig. 2.5C**). Specifically, a one-way Kruskal-Wallis test revealed a significant difference in response latency among the three classes ($\chi^2_{(2, 1483)} = 42.5$, $p < .001$), and post-hoc Dunn tests revealed that both the adapting and

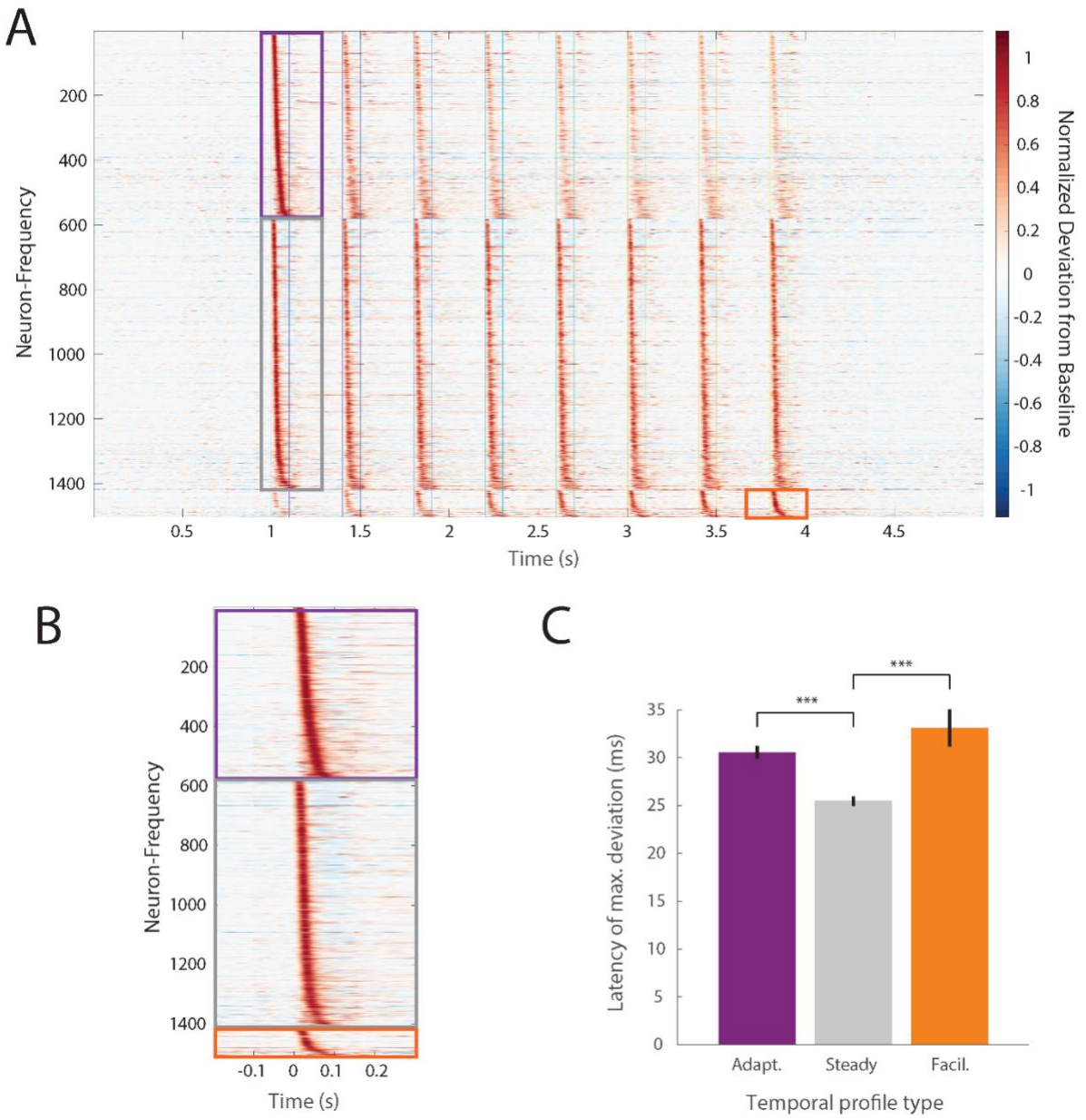

Figure 2.5 Evoked spike timing differs between temporal profile groups according to model predictions.

A. Image plot of each neuron-frequency pair's PSTH, clustered by temporal profile (adapting, steady, facilitating) and sorted within each temporal profile by the latency of maximal evoked firing. Each row represents a single unit's average PSTH, expressed as the normalized deviation from baseline firing rate. Within adapting and steady temporal profiles, units are sorted by latency of maximum deviation to the 1st tone. Within the facilitating temporal profile, units are sorted by latency of maximum deviation to the 8th tone. **B**. The same data in A re-plotted (see outlines) to show the distribution of average spike timing across trials for the time regions surrounding the peak response of that temporal profile. **C**. Bar plot comparing the mean tone-evoked latencies within each temporal profile. Error lines indicate the standard error of the mean. There was a significant difference in response latency among the three classes (χ^2 (2, 1483) = 42.5, $p < .001$). Both the adapting and facilitating classes of neuron-frequency pairs had significantly longer mean response latencies than the stable class (adapting vs. stable, $Z = 6.18$, $p < .001$; facilitating vs. stable, $Z = 3.25$, $p < .001$), but the adapting and facilitating classes were not different from each other (adapting vs. facilitating, $Z = -0.28$, $p = .29$).

facilitating classes of neuron-frequency pairs had significantly longer mean response latencies than the stable class (adapting vs. stable, Z = 6.18, $p < .001$; facilitating vs. stable, Z = 3.25, $p < .001$). However, the adapting and facilitating classes were not different from each other (adapting vs. facilitating, Z = -0.28, $p = .29$) (**Fig. 2.5C**). This result provided support for the notion that pyramidal cells that fire with a short latency — presumably because they receive strong excitatory synaptic inputs — undergo less temporal-context modulation because dynamic changes in IPSP strength are less effective in influencing pyramidal neurons spiking generated by feed-forward activity. In other words, the inhibitory inputs onto these units are functionally weak because they are delayed relative to excitation.

Finally, because there was a large degree of variation in the response latency *within* a temporal profile, we considered whether the degree of adaptation or facilitation for a given unit-frequency was correlated with its response latency. We quantified the degree of adaptation or facilitation as the *t*-statistic resulting from the linear regression of tone-evoked spikes by the tone's serial position (see Methods). We found that for adapting unit-frequencies the degree of adaptation was mildly correlated with response latency ($r = -.194$, $p < .001$) such that more strongly adapting unit-frequencies had longer evoked latencies. For facilitating unit-frequencies, this relationship was not significant.

2.2.6 Model predicts a paradoxical decrease in firing caused by prior PV inactivation.

A long-standing challenge has been to establish a causal role for STP in temporal-context modulation. Interestingly, because of the model's reliance on STP to

account for temporal-context effects, it led to a unique and counterintuitive prediction. Specifically, because by definition STP is dependent on the previous spike history on the time scale of tens-to-hundreds of milliseconds, preventing spikes evoked by a tone in one class of neurons should prevent STP and alter responses to subsequent tones. Indeed, as shown in **Figure 2.6A,** optogenetic inactivation of PV neurons (due to hyperpolarization driven by light activation of Cre-dependent ArchT) can sometimes not produce any effect during the optogenetically inhibited tone, but can alter the Pyr neuron response to a tone presented 400 ms later.

We first simulated optogenetic inhibition of PV and SST neurons by strongly hyperpolarizing either PV or SST model units from 100 ms before to 150 ms after the first tone presentation. Inhibitory inactivation of PV, but not SST, unit caused a history-dependent effect on Pyr unit spiking under stable and facilitating weight regimes (**Fig. 2.6B**). More specifically, in the stable weight regime, inactivating the PV unit during the first tone caused a *decrease* in the number of Pyr spikes evoked by the second tone (sign rank Z = -3.59; $p < .001$). We will refer to this as an "n+1" effect because manipulation during one tone alters the response to the next tone in the absence of any further manipulation. A similar effect of slightly smaller magnitude was observed for the facilitating weight regime (sign rank Z = -3.42; $p < .001$). The n+1 effect was primarily a result of the STP profile of the PV→Pyr synapse (**Fig. 2.2B**). Because the PV unit normally spikes robustly during the first tone presentation, the PV→Pyr synapse is normally weakened during the second tone presentation due to its strong short-term depression—this depression normally counterbalances some of the some of the short-term depression of the EPSPs generated by the inputs. Hyperpolarizing and thus

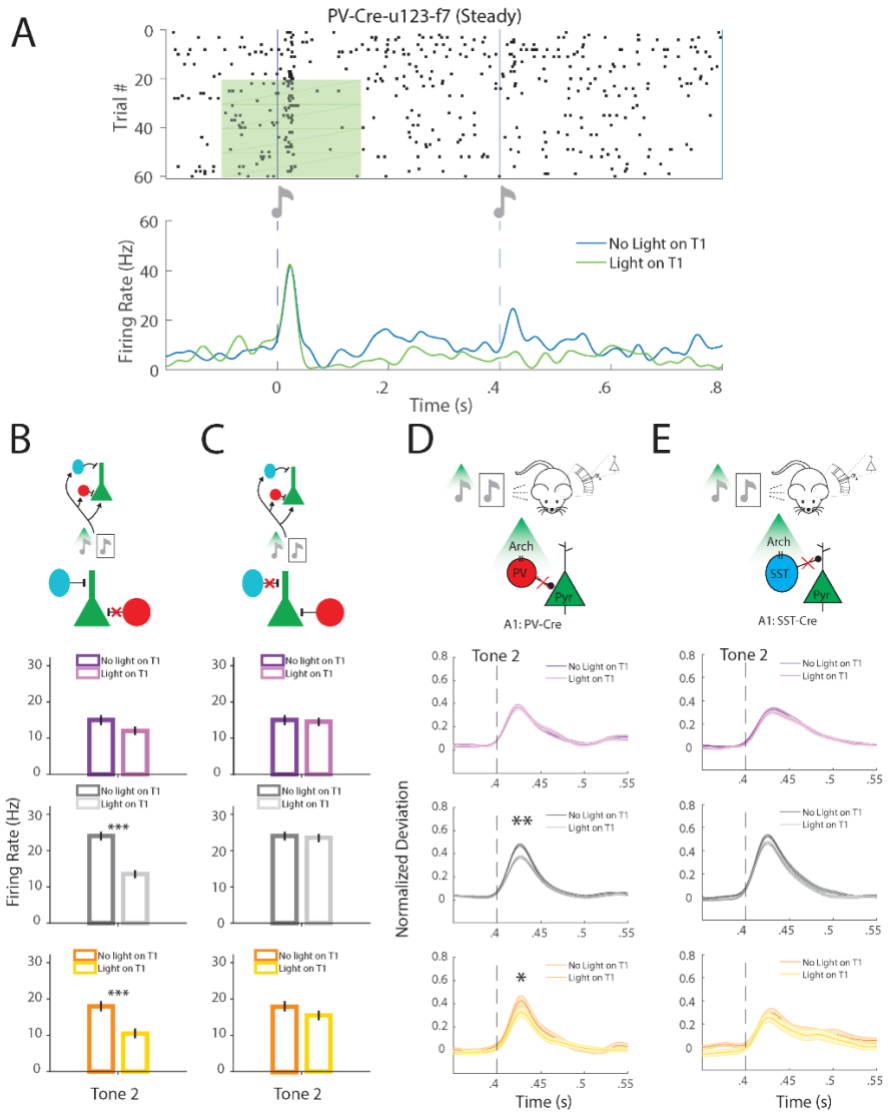

Figure 2.6 Simulated inactivation of inhibitory units correctly predicts that PV interneuron inactivation during the 1st tone causes a decrease in the tone-evoked firing rate during the 2nd tone for steady and facilitating units.

A. Example of an experimentally recorded Pyr neuron in which inactivation of PV interneurons during the 1st tone decreased the firing rate evoked by the 2nd tone. In the spike raster (upper), the green rectangle indicates trials and window of optical stimulation. The line plots (lower) display PSTHs separately for trials with and without optical stimulation during the 1st tone. **B**. Effect of simulated PV inhibition on Pyr unit firing rates in adapting (purple), steady (grey), or facilitating (orange) regimes. Bars indicate the mean firing rate evoked by the 2nd in a simulated pair of "tones" with and without simulated optogenetic inactivation of the PV unit during the 1st tone. Means were taken across 20 trials. Significant differences were found for the steady and facilitating weight regimes.**C**. Simulated SST inhibition on Pyr unit firing rate. SST inhibition during the 1st tone did not significantly alter Pyr firing in any of the three regimes ($p > .05$). **D**. Experimentally observed effects of PV inhibition on Pyr firing to the subsequent tone, across all three temporal profile classes. Shaded regions indicate the standard error of the mean. Significant differences were found only for the steady and facilitating neuron-frequency pairs. **E**. Mean normalized evoked firing rates of adapting (purple), steady (grey), or facilitating (orange) neuron-frequency pairs in response to SST inhibition. Consistent with the model predictions of the model no significant differences were observed in either class of temporal profiles ($p > .05$). *p < 0.05, **p < 0.01, ***p < 0.001

preventing the PV unit from spiking during the first tone effectively postpones its initial and strongest inhibitory effect until the second tone. Thus, following PV inactivation on the first tone, PV inhibition is stronger on the second tone, reducing the Pyr unit firing rate. The n+1 effect in response to PV neuron inactivation was not observed in the adapting weight regime ($p > 0.05$), because it is dominated by SST inhibition (**Fig. 2.6B,** top panel). In contrast to PV inactivation, no effect of prior inactivation was observed for SST inactivation (all $p > 0.05$; **Fig. 2.6C**), both because the SST→Pyr synapse exhibited relatively little STP and because the SST unit is relatively inactive during the first tone presentation.

Finally, we tested whether the n+1 effect was present *in vivo* (**Fig. 2.6D-E**). We contrasted the trials on which there was light inactivation during the first tone to trials in which there was no light inactivation during the first tone, and compared the number of spikes evoked by the second tone (note that there was no optogenetic inactivation during the second tone). In PV-Cre mice, there was a significant n+1 effect for stable neuron-frequency pairs (sign rank $Z = -3.10$; $p = .002$) and a significant but weaker prior light effect for facilitating neuron-frequency pairs (sign rank $Z = -2.29$; $p = .02$). **Fig. 2.6E** shows a raster plot and PSTH of a sample neuron that exhibited the n+1 effect—specifically inhibition of PV neurons during the first tone increased the number of spikes in response to the second tone. As in the computational model, the n+1 effect was not observed in the adapting group in response to PV inhibition ($p > .05$). Again, as in the model, there was no n+1 effect produced by SST inactivation in either the adapting, stable, or facilitating groups (all $p > .05$). These results provide a strong validation of a

counterintuitive prediction of the model, and strong support for the role of STP in temporal-context modulation.

2.3 Discussion

A large body of experimental evidence has demonstrated that there is significant diversity in the temporal profile of the firing property of cortical neurons in response to repeated stimuli. However, the origin this diversity has remained largely unexplored. The current study identified a novel mechanism underlying temporal-context modulation of sensory stimuli as the source of the diversity of temporal response profiles. We found that a simple feedforward model that incorporated empirically-based STP properties at five different synapses can account for the three experimentally observed temporal profiles by simply altering the relative weights of PV and SST inhibition. Critically the model generated two predictions that were experimentally tested and validated.

The first prediction was a relationship between temporal-profile class and tone-evoked spike latency. This prediction arises as a consequence of STP at the inhibitory branches the circuit (i.e., Input→PV→Pyr and Input→SST→Pyr), and, that for this STP to modulate pyramidal firing the EPSPs and IPSPs onto a pyramidal neuron must overlap in time. Thus, the strong temporal-context modulation in the adapting and facilitating groups suggest a broad temporal overlap between excitation and inhibition. In contrast, the relatively weak temporal context modulation in the stable populations, suggests that these neurons may fire too quickly to be strongly modulated by the inhibitory branch of the circuit which is delayed as a result of the additional synaptic step. This prediction was confirmed by demonstrating that the latency of the stable

population was significantly shorter than that of the adapting and facilitating population (**Fig. 2.5**).

The second prediction of the model was that because temporal-context modulation is driven by changes in E/I balance imposed by STP, that "blocking" STP should result in what we refer to as an "n+1" effect. Specifically, that optogenetically inhibiting PV neurons should prevent potential STP at the PV→Pyr synapse. If this STP is contributing to temporal-context modulation, then inhibition during tone *n* should influence the pyramidal neuron response at tone *n+1*. This novel prediction was experimentally confirmed by showing that inactivation of PV neurons at the first tone decreased firing of the pyramidal neurons in response to the second tone (**Fig. 2.6**). In contrast, this effect was not observed by inhibiting the SST neurons, a result also predicted by the model because unlike the significant short-term depression at the PV→Pyr, the SST→Pyr IPSP exhibits weak and inconsistent forms of STP. This result provides some of the best evidence to date that STP contributes to sensory adaptation.

2.3.1 STP generates temporal-context modulation through spike rate and latency shifts

It is increasingly recognized that PV and SST neurons fulfill distinct computational roles within cortical microcircuits [e.g., 6, 22, 26, 52, 53, 54]. As both PV and SST neurons inhibit pyramidal neurons, a crucial question pertains to which differences between PV and SST neurons underlie their different computational roles. Critical differences include distinct connectivity patterns (e.g., SST neurons primarily target pyramidal dendrites), and intrinsic excitability (PV neurons are fast-spiking, and SST are low-threshold)[42, 43]. But one of the most robust differences between PV and

SST neurons lies in their STP properties, specifically, excitatory inputs to PV neurons generally undergo short-term synaptic depression, while the excitatory synapses onto SST neurons exhibit dramatic facilitation [38, 39, 41, 44, 55]. Using computational and experimental approaches we were able to quantify the contributions of inhibition not only to adaptation, but also to facilitation, and distinguish between firing rate and spike latency contributions. Our results provide for a novel differential role for PV and SST in temporal context modulation. Further, we find that the diversity of temporal response profiles in cortical neurons may be explained by the initial connectivity of local circuits—i.e., whether the dominant source of inhibition onto a pyramidal neuron is tilted towards PV or SST neurons.

In contrast to previous models of how STP may underlie temporal context modulation, which have relied on abstract or firing rate implementations, our spiking model emphasizes a key functional mechanism of STP. The dynamic decreases and increases in synaptic strength imposed by STP are not expressed solely by changing spike probability and rate, but by relatively small shifts in the latency of inhibitory neuron firing. That is, in the temporal relationship between the EPSPs and IPSPs impinging on the pyramidal neurons. Indeed, previous experimental and computational results have demonstrated that both short- and long-term synaptic plasticity can shape neural computations by modulating the race to threshold between EPSPs and IPSPs. For example, latency delays of a mere millisecond in inhibitory firing can convert a subthreshold excitatory input in to a suprathreshold input [48, 49, 51]. Thus, the balance between excitation and inhibition is not simply governed by absolute synaptic strength but by the relative timing of EPSPs and IPSPs in the pyramidal neurons. Thus a further

prediction that emerges from the model is that significant shifts in the latency of SST neurons should shorten in response to consecutive tones. And that it is possible that short-term facilitation onto SST neurons might not significantly increase SST firing rate, but still underlie adaptation through shifts in latency.

2.3.2 Limitations of the model

The model presented here is highly simplified in that it does not take into account the full complexity of cortical microcircuits, including the presence of recurrent excitation, and the connections between SST→PV and PV→SST neurons [56, 57]. Additionally, we focused on a single neurons, rather than a large-scale population model [58]. These simplifying assumptions allowed us to firmly ground the model in the empirical data and avoid making assumptions about the recurrent circuitry underlying cross-frequency interaction. Importantly, it is unlikely that that these higher-order circuit interactions significantly contribute to the short-latency pyramidal neuron responses to single tones being studied here. In order to further understand the cross-frequency suppression and facilitation effects, in the future, detailed population-level models will have to be developed. Towards this goal the current results highlight the importance of empirically defining the detailed microcircuit structure as latency differences of a few milliseconds can have a profound functional effect.

Critical to our objective of anchoring the model in the experimental data, are the assumptions related to the direction, magnitude, and temporal profile of STP at each synapse type. The model is largely robust to the changes in the empirically-derived values regarding the magnitude and times constants of STP. However, the presence of short-term facilitation at the Input→SST synapse, and depression at the PV→Pyr

synapses are key to the presented results. And while there is significant variability in the experimental estimated of the STP parameters at the synapse classes simulated here, to the best of our knowledge all experimental data points to facilitation and depression of Input→SST and PV→Pyr synapses respectively [38, 39, 41, 44, 55, 59]. Similarly, experimentally observed differences in the intrinsic properties of PV and SST neurons were also incorporated into the model. These included the time constants, spike threshold, spike adaptation, and the localization of inhibitory inputs onto the Pyr units. While these properties influence model behavior, they are not fundamental to obtaining the reported results.

2.3.3 Predictions and Conclusions

In addition to the predictions tested here a number of additional predictions emerge from the current study, including: 1) the first-spike latency of SST neurons should decrease in response to consecutive tones; 2) adapting and facilitating neurons should receive strong SST and PV inhibition, respectively; and 3) that experience-dependent increases in responses to tone sequences [21, 60, 61] might be attributable to SST plasticity—e.g., long-term depression of SST IPSPs. Overall our results suggest that differential PV and SST inhibition contributes to the diversity of temporal context profiles [6, 22, 26]. Importantly the current results provide compelling evidence that short-term synaptic plasticity is indeed a key mechanism driving temporal context modulation, by characterizing a novel *n+1* effect that arises as a direct consequence of STP.

2.4 Methods

2.4.1 In vivo electrophysiology

The experimental data set used here was the same as that used in a previously published data set [6]. Eleven adult male PV-Cre (B6;129P2-Pvalbtm1(cre)Arbr/J) and SST-Cre (Ssttm2.1(cre)Zjh/J) mice aged 12-15 weeks were anesthetized, and AAV encoding Cre-dependent ArchT [6] were injected into auditory cortex (A1) 2-4 weeks prior to experimental recordings. For the electrophysiology recording session mice were again anesthetized, and a linear silicon 32-electrode probe placed into A1. Online monitoring of sound-evoked responses was used to ensure correct electrode positioning and to identify a 1.3 octave range of pure tone frequencies between 1 and 80 kHz that evoked multi-unit responses ranging from maximal to weak. Offline spike sorting was performed using commercial software.

Acoustic tone stimuli were 100-ms tone pips separated by 300 ms of silence (400 ms stimulus-onset-asynchrony). Trains of eight tones of a single frequency were presented, separated by 2.4 s of silence. Each train used one of 10 frequencies within the 1.3 octave range in a pseudorandom and counterbalanced manner. In half of the trials, an optic fiber was used to direct 532-nm laser light into A1 from 100 ms before to 150 ms after each of 4 time periods: during the first tone, during the last tone, or during the silent period 400 ms before or 400 ms after the train.

2.4.2 Analysis of A1 recordings

Temporal profiles to the presentation of eight tones of the same frequency were classified as adapting, stable, or facilitating, by regressing the number of tone-evoked spikes against the corresponding serial position. Statistical tests (sign rank, rank sum,

or linear regression, see below) were performed using single-trial spike counts within the 10-70 ms period following tone onset as the outcome variable, with the significance level α set to 0.05. *Adapting* temporal profiles were defined by a significant evoked response to the first tone *and* a significant negative slope across serial positions. *Facilitating* profiles were defined by a significant evoked response to the final tone and a significant positive regression slope. *Steady* temporal profiles were defined as those with a significant response to both the first and final tones of a train and a nonsignificant regression. As in previous studies, analyses were based on neuron-frequency pairs, i.e., each of the 10 frequencies used in the tone trains was tested and analyzed separately for each neuron [6].

To help ensure that our analyses only included pyramidal cells, we took advantage of the Cre-dependent ArchT expression in the following manner: For PV-Cre mice, adapting and steady unit-frequencies were excluded if they had significantly lower 1^{st}-tone evoked firing rate during light compared to non-light. For SST-Cre mice, steady and facilitating unit-frequencies were excluded if they had significantly lower 8^{th}-tone evoked firing rate during light compared to non-light. Although indirect light-driven decreases in activity in pyramidal cells could be produced through disinhibition (e.g. in a SOM → PV → Pyr) this method decreases the likelihood of false-positive Pyr units in our analyses. Furthermore, the number of units that exhibited a light-driven decrease in activity was small: 22 adapting unit-frequencies (3.8%), 30 steady unit-frequencies (3.6%), and 3 facilitating unit-frequencies (3.5%) were excluded in this manner.

Post-stimulus time histograms (PSTHs) were calculated by convolving spike times with a Gaussian kernel of standard deviation 10 ms. When averaged across

neuron-frequencies, PSTHs for each neuron-frequency were converted to a normalized deviation from baseline firing rate in the following manner: First, we calculated the mean and standard deviation of the firing rate in the baseline period during the 1 second period between trials. Next, we subtracted the mean from the PSTH and divided by the standard deviation of the baseline firing rate. Finally, we normalized each PSTH such that its maximum was 1.

For the latency analysis (**Fig. 2.5**), PSTHs were calculated in the standard manner by summing spikes within 4 ms time bins. A neuron-frequency's tone-evoked latency was operationalized as the center of the time bin with the highest firing rate between 10 and 100 ms following tone onset for either the first (adapting and steady profiles) or last tone (facilitating profiles). To test for the effect of prior light on the evoked firing rate at serial position 2 (**Fig. 2.6**), sign rank tests were employed.

2.4.3 Computational Model

Units were simulated as conductance-based integrate-and-fire units (**Table 2.1**). Three distinct types of neuronal units were implemented: excitatory pyramidal (Pyr), fast-spiking inhibitory Parvalbumin (PV), and low-threshold inhibitory Somatostatin (SST) units. The three unit types were modeled with distinct intrinsic parameters including membrane time constant, spike threshold, and an afterhyperpolarization current designed to reproduce their firing properties and spike-adaptation (see **Table 2.2**). Pyr units were modeled as two-compartment (soma and dendrite) units, while both types of inhibitory units were simulated as a single compartment unit. Simulations were implemented in the NEURON simulation environment [62], and were based on previously published models [49, 63]. Each unit's membrane potential was subject to

Table 2.1 Global Parameters in Chapter 2.

Parameter	Value
Integration time step (ms)	0.05
Temperature (°C)	36
Cell Length (cm)	10
Cell Diameter (cm)	10
Resting membrane potential (mV)	-65
Leak conductance (uS / cm^2)	100
Internal resistance (Ω·cm)	35
Refractory period (ms)	3
AHP reversal potential (mV)	-80
Dendritic compartment length (cm)	100
Dendritic compartment diameter (cm)	0.5
Noise Current (nA)	0.005

Global NEURON parameters used for all units and simulations.

Table 2.2 Unit Parameters in Chapter 2.

Parameter	Pyr	PV	SST
Membrane Time Constant (ms)	15	7.5	19
Spike threshold (mV)	-35	-40	-45
Spike duration (ms)	1	0.5	0.75
Membrane capacitance (μF/cm^2)	1.5	0.75	1.9
AHP Conductance Incremement (uS / cm^2 / spike)	100	25	50
AHP Decay Time Constant (ms)	5	1	3

NEURON parameters that differed between units.

leak, afterhyperpolarization, and synaptic currents as follows:

$$C_m \frac{dV}{dt} = g_L(V_m - E_L) + g_{AHP}(V_m - E_{AHP}) + g_{syn}(V_m - E_{syn})$$

Excitatory and inhibitory synaptic transmission was modeled using Alpha synapses with forward and backward binding [64]. Short-term synaptic plasticity (STP) was incorporated at all synapses using the Tsodyks-Markram formulation [65, 66], wherein repeated presynaptic spikes modulated the maximal synaptic conductance by a product of resources (R) and availability (u):

$$R_{n+1} = 1 - [1 - R_n(1 - u_n)] \cdot e^{-\frac{\Delta t_{sp}}{\tau_d}}$$

$$u_{n+1} = U + u_n(1 - U) \cdot e^{-\frac{\Delta t_{sp}}{\tau_f}}$$

where τ_d and τ_f are the time constants of depression and facilitation respectively, and U can be interpreted as the initial release probability. R and u are updated at the time of each presynaptic spike and Δt_{sp} is the interval between the current and previous spike. All cellular and synaptic parameters were defined in consultation with the NeuroElectro project and studies of paired patch-clamp recordings from connected cortical cells [36, 38-41, 55, 67]. Synaptic parameters are shown in **Table 2.3**.

Each unit received 3 distinct input fibers, and tones were simulated as a single spike in each input fiber once in sequence with a fixed 5 ms inter-spike interval (simulations with inter-spike intervals in the 5-10 ms range and the same weight parameters yielded similar results). Optogenetic inactivation was modeled by transiently injecting a current of -.01 nA to the inactivated unit. The simulated circuit was based on the feedforward connectivity of a cortical microcircuit [53, 68], and thus did not incorporate the positive feedback between pyramidal neurons, or negative feedback

Table 2.3 Synaptic Parameters in Chapter 2.

Parameter	Inp-Pyr	Inp-PV	Inp-SST	PV-Pyr	SST-Pyr
Synaptic Delay (ms)	3	1	2	0.5	0.5
U	0.15	0.25	0.05	0.35	0.2
τ_d (ms)	250	500	10	800	100
τ_f (ms)	10	10	800	10	100

NEURON parameters that differed between synapses.

produced by pyramidal to inhibitory neuron activation. This simplifying assumption was made because we focused primarily on the short-latency responses of A1 neurons which are less likely to be influenced by feedback circuitry [7]. Furthermore, as we simulated a reduced circuit, the relatively large single synaptic currents reflect the near synchronous activation of multiple synapses [69, 70]. This simplified circuit allowed a more tractable approach that was better constrained by the experimental data.

2.4.4 Experimental Design and Statistical Analysis

For our analysis of the experimental dataset considered here [6], between-subjects variables included mouse genotype (PV-Cre or SST-Cre) and the temporal profile (adapting, steady, or facilitating). From 5 male PV-Cre mice, there were 225 adapting, 413 steady, and 45 facilitating unit-frequencies. From 6 male SST-Cre mice, there were 350 adapting, 412 steady, and 41 facilitating unit-frequencies. Within-subjects variables included the serial position within each train of tone pips (1-8) and the presence or absence of light during optogenetic inactivation experiments (non-light vs. light). To classify unit-frequencies based on their temporal profile (**Fig. 2.1**), and to

assess firing latency (**Fig. 2.5C**), epochs containing light stimulation were excluded, yielding a minimum of 40 non-light epochs for each unit-frequency and serial position. Separating trials to assess the effect of prior light on the evoked activity at serial position 2 (**Fig. 2.6**) yielded a minimum of 20 epochs per condition. To classify unit-frequencies based on their temporal profile, Wilcoxon signed rank tests (to test for the presence of evoked activity), linear regression models (to determine the slope of the temporal profile), and Wilcoxon rank sum tests (to exclude potential interneurons) were employed. To compare latencies between temporal profiles (**Fig. 2.5**), a one-way Kruskal-Wallis test was used, and the Dunn procedure was used to perform pairwise comparisons between the three temporal profiles. To test for the effect of prior light at tone 2 (**Fig. 2.6**), Wilcoxon signed rank tests were employed.

For statistical analysis of the simulation results, we used the same statistical tests as for the experimental data. The between-subjects variable was the interneuron unit type that received simulated optogenetic inactivation (PV or SST). Within-subjects variables included the serial position within each train of tones (1-8) and presence or absence of simulated inactivation. For all simulations that were subject to statistical tests, 20 trials were performed for each condition.

2.5 Chapter 2 References

1. Doupe, A.J. and P.K. Kuhl, *Birdsong and human speech: common themes and mechanisms.* Annu Rev Neurosci, 1999. **22**: p. 567-631.

2. Tallal, P., *Improving language and literacy is a matter of time.* Nat Rev Neurosci, 2004. **5**(9): p. 721-8.

3. Saberi, K. and D.R. Perrott, *Cognitive restoration of reversed speech.* Nature, 1999. **398**(6730): p. 760-760.

4. Nelken, I. and G. Chechik, *Information theory in auditory research.* Hearing Research, 2007. **In Press, Corrected Proof**.

5. Brosch, M. and C.E. Schreiner, *Time course of forward masking tuning curves in cat primary auditory cortex.* J Neurophysiol, 1997. **77**(2): p. 923-43.

6. Natan, R.G., W. Rao, and M.N. Geffen, *Cortical Interneurons Differentially Shape Frequency Tuning following Adaptation.* Cell Reports, 2017. **21**(4): p. 878-890.

7. Rennaker, R.L., et al., *Anesthesia suppresses nonsynchronous responses to repetitive broadband stimuli.* Neuroscience, 2007. **145**(1): p. 357-369.

8. Phillips, D.P., S.E. Hall, and J.L. Hollett, *Repetition rate and signal level effects on neuronal responses to brief tone pulses in cat auditory cortex.* The Journal of the Acoustical Society of America, 1989. **85**(6): p. 2537-2549.

9. Shu, Z.J., N.V. Swindale, and M.S. Cynader, *Spectral motion produces an auditory after-effect.* Nature, 1993. **364**(6439): p. 721-723.

10. Klampfl, S., et al., *A quantitative analysis of information about past and present stimuli encoded by spikes of A1 neurons.* Journal of Neurophysiology, 2012. **108**(5): p. 1366-1380.

11. Brosch, M. and C.E. Schreiner, *Sequence Sensitivity of Neurons in Cat Primary Auditory Cortex.* Cereb. Cortex, 2000. **10**(12): p. 1155-1167.

12. Wehr, M. and A.M. Zador, *Synaptic mechanisms of foward suppression in rat auditory cortex.* Neuron, 2005. **47**: p. 437-445.

13. Sadagopan, S. and X. Wang, *Nonlinear Spectrotemporal Interactions Underlying Selectivity for Complex Sounds in Auditory Cortex.* J. Neurosci., 2009. **29**(36): p. 11192-11202.

14. Katz, Y., J.E. Heiss, and I. Lampl, *Cross-Whisker Adaptation of Neurons in the Rat Barrel Cortex.* J. Neurosci., 2006. **26**(51): p. 13363-13372.

15. Ahissar, E., et al., *Temporal frequency of whisker movement. II. Laminar organization of cortical representations.* J Neurophysiol, 2001. **86**(1): p. 354-67.

16. Chung, S., X. Li, and S.B. Nelson, *Short-term depression at thalamocortical synapses contributes to rapid adaptation of cortical sensory responses in vivo.* Neuron, 2002. **34**(3): p. 437-446.

17. Ohzawa, I., G. Sclar, and R.D. Freeman, *Contrast gain control in the cat visual cortex.* Nature, 1982. **298**(5871): p. 266-268.

18. Wilson, D.A., *Odor specificity of habituation in the rat anterior piriform cortex.* Journal of Neurophysiology, 2000. **83**(1): p. 139-145.

19. Kohn, A., *Visual adaptation: physiology, mechanisms, and functional benefits.* J Neurophysiol, 2007. **97**(5): p. 3155-64.

20. Gutnisky, D.A. and V. Dragoi, *Adaptive coding of visual information in neural populations.* Nature, 2008. **452**(7184): p. 220-4.

21. Kilgard, M.P. and M.M. Merzenich, *Order-sensitive plasticity in adult primary auditory cortex.* Proc Natl Acad Sci U S A, 2002. **99**(5): p. 3205-9.

22. Phillips, E.A.K., C.E. Schreiner, and A.R. Hasenstaub, *Cortical Interneurons Differentially Regulate the Effects of Acoustic Context.* Cell Reports, 2017. **20**(4): p. 771-778.

23. Solomon, Samuel G. and A. Kohn, *Moving Sensory Adaptation beyond Suppressive Effects in Single Neurons.* Current Biology, 2014. **24**(20): p. R1012-R1022.

24. Rose, G., C. Leary, and C. Edwards, *Interval-counting neurons in the anuran auditory midbrain: factors underlying diversity of interval tuning.* Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 2011. **197**(1): p. 97-108.

25. Klump, G.M. and H.C. Gerhardt, *Use of non-arbitrary acoustic criteria in mate choice by female gray tree frogs.* Nature, 1987. **326**: p. 286–88.

26. Natan, R.G., et al., *Complementary control of sensory adaptation by two types of cortical interneurons.* eLife, 2015. **4**: p. e09868.

27. Phillips, E.A.K., C.E. Schreiner, and A.R. Hasenstaub, *Diverse effects of stimulus history in waking mouse auditory cortex.* Journal of Neurophysiology, 2017. **118**(2): p. 1376-1393.

28. Rosenbaum, R., J. Rubin, and B. Doiron, *Short Term Synaptic Depression Imposes a Frequency Dependent Filter on Synaptic Information Transfer.* PLOS Computational Biology, 2012. **8**(6): p. e1002557.

29. Buonomano, D.V., *Decoding temporal information: a model based on short-term synaptic plasticity.* J Neurosci., 2000. **20**: p. 1129-1141.

30. Fortune, E.S. and G.J. Rose, *Short-term synaptic plasticity as a temporal filter.* Trends Neurosci, 2001. **24**(7): p. 381-5.

31. Motanis, H., M.J. Seay, and D.V. Buonomano, *Short-Term Synaptic Plasticity as a Mechanism for Sensory Timing.* Trends in Neurosciences, 2018. **41**(10): p. 701-711.

32. Oswald, A.-M.M. and A.D. Reyes, *Maturation of Intrinsic and Synaptic Properties of Layer 2/3 Pyramidal Neurons in Mouse Auditory Cortex.* J Neurophysiol, 2008. **99**(6): p. 2998-3008.

33. Abbott, L.F. and W.G. Regehr, *Synaptic computation.* Nature, 2004. **431**(7010): p. 796-803.

34. Buonomano, D.V. and M.M. Merzenich, *Temporal information transformed into a spatial code by a neural network with realistic properties.* Science, 1995. **267**: p. 1028–30.

35. Pérez, O. and H. Merchant, *The Synaptic Properties of Cells Define the Hallmarks of Interval Timing in a Recurrent Neural Network.* The Journal of Neuroscience, 2018. **38**: p. 4186-4199.

36. Levy, R.B. and A.D. Reyes, *Coexistence of Lateral and Co-Tuned Inhibitory Configurations in Cortical Networks.* PLOS Computational Biology, 2011. **7**(10): p. e1002161.

37. Xu, H., V.C. Kotak, and D.H. Sanes, *Conductive Hearing Loss Disrupts Synaptic and Spike Adaptation in Developing Auditory Cortex.* Journal of Neuroscience, 2007. **27**(35): p. 9417-9426.

38. Beierlein, M., J.R. Gibson, and B.W. Connors, *Two dynamically distinct inhibitory networks in layer 4 of the neocortex.* J Neurophysiol, 2003. **90**(5): p. 2987-3000.

39. Takesian, A.E., et al., *Hearing loss differentially affects thalamic drive to two cortical interneuron subtypes.* Journal of Neurophysiology, 2013. **110**(4): p. 999-1008.

40. Levy, R.B. and A.D. Reyes, *Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex.* J Neurosci, 2012. **32**(16): p. 5609-19.

41. Tan, Z., et al., *Robust but delayed thalamocortical activation of dendritic-targeting inhibitory interneurons.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(6): p. 2187-2192.

42. Gupta, A., Y. Wang, and H. Markram, *Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex.* Science, 2000. **287**(5451): p. 273-8.

43. Oswald, A.M.M. and A.D. Reyes, *Development of inhibitory timescales in auditory cortex.* Cerebral Cortex, 2011. **21**(6): p. 1351-1361.

44. Silberberg, G. and H. Markram, *Disynaptic Inhibition between Neocortical Pyramidal Cells Mediated by Martinotti Cells.* Neuron, 2007(53): p. 735-746.

45. Takesian, A.E., V.C. Kotak, and D.H. Sanes, *Presynaptic GABAB Receptors Regulate Experience-Dependent Development of Inhibitory Short-Term Plasticity.* Journal of Neuroscience, 2010. **30**(7): p. 2716-2727.

46. Gibson, J.R., M. Beierlein, and B.W. Connors, *Two networks of electrically coupled inhibitory neurons in neocortex.* Nature, 1999. **402**(6757): p. 75-79.

47. Tripathy, S.J., et al., *NeuroElectro: a window to the world's neuron electrophysiology data.* Frontiers in Neuroinformatics, 2014. **8**(April): p. 1-11.

48. Marder, C.P. and D.V. Buonomano, *Timing and balance of inhibition enhance the effect of long-term potentiation on cell firing.* J Neurosci, 2004. **24**(40): p. 8873-84.

49. Carvalho, T.P. and D.V. Buonomano, *Differential effects of excitatory and inhibitory plasticity on synaptically driven neuronal input-output functions.* Neuron, 2009. **61**(5): p. 774-85.

50. Pouille, F. and M. Scanziani, *Enforcement of temporal fidelity in pyramidal cells by somatic feed-forward inhibition.* Science, 2001. **293**: p. 1159-1163.

51. Wilent, W.B. and D. Contreras, *Dynamics of excitation and inhibition underlying stimulus selectivity in rat somatosensory cortex.* Nat Neurosci, 2005. **8**(10): p. 1364-70.

52. Atallah, B.V., et al., *Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli.* Neuron, 2012. **73**(1): p. 159-70.

53. Cardin, J.A., *Inhibitory Interneurons Regulate Temporal Precision and Correlations in Cortical Circuits.* Trends in Neurosciences, 2018. **41**(10): p. 689-700.

54. Kato, H.K., S.K. Asinof, and J.S. Isaacson, *Network-Level Control of Frequency Tuning in Auditory Cortex.* Neuron, 2017. **95**(2): p. 412-423.e4.

55. Reyes, A., et al., *Target-cell-specific facilitation and depression in neocortical circuits.* Nat Neurosci, 1998. **1**(4): p. 279-285.

56. Walker, F., et al., *Parvalbumin- and vasoactive intestinal polypeptide-expressing neocortical interneurons impose differential inhibition on Martinotti cells.* Nature Communications, 2016. **7**(1): p. 13664.

57. Pfeffer, C.K., et al., *Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons.* Nat Neurosci, 2013. **16**(8): p. 1068-1076.

58. Park, Y. and M.N. Geffen, *A circuit model of auditory cortex.* PLOS Computational Biology, 2020. **16**(7): p. e1008016.

59. Pala, A. and Carl C.H. Petersen, *In Vivo Measurement of Cell-Type-Specific Synaptic Connectivity and Synaptic Transmission in Layer 2/3 Mouse Barrel Cortex.* Neuron, 2015. **85**(1): p. 68-75.

60. Zhou, X., et al., *Successive-signal biasing for a learned sound sequence.* Proceedings of the National Academy of Sciences, 2010. **107**(33): p. 14839-14844.

61. Engineer, C.T., et al., *Cortical activity patterns predict speech discrimination ability.* Nat Neurosci, 2008. **11**: p. 603-608.

62. Hines, M.L. and N.T. Carnevale, *The NEURON simulation environment.* Neural Comput, 1997. **9**: p. 1179-1209.

63. Goudar, V. and D.V. Buonomano, *A model of order-selectivity based on dynamic changes in the balance of excitation and inhibition produced by short-term synaptic plasticity.* Journal of Neurophysiology, 2015. **113**(2): p. 509-523.

64. Destexhe, A., Z.F. Mainen, and T.J. Sejnowski, *An efficient method for computing synaptic conductances based on a kinetic model of receptor binding.* Neural Comput., 1994. **6**: p. 14-18.

65. Markram, H., Y. Wang, and M. Tsodyks, *Differential signaling via the same axon of neocortical pyramidal neurons.* Proc. Natl. Acad. Sci. USA., 1998. **95**: p. 5323-5328.

66. Carvalho, T.P. and D.V. Buonomano, *A novel learning rule for long-term plasticity of short-term synaptic plasticity enhances temporal processing.* Front Integr Neurosci, 2011. **5**: p. 20.

67. Reyes, A. and B. Sakmann, *Developmental switch in the short-term modification of unitary EPSPs evoked in layer 2/3/ and layer 5 pyramidal neurons of rat neocortex.* J Neurosci., 1999. **19**: p. 3827-3835.

68. Douglas, R.J. and K.A.C. Martin, *Mapping the Matrix: The Ways of Neocortex.* Neuron, 2007. **56**(2): p. 226-238.

69. Bruno, R.M. and B. Sakmann, *Cortex Is Driven by Weak but Synchronously Active Thalamocortical Synapses.* Science, 2006. **312**(5780): p. 1622-1627.

70. DeWeese, M.R. and A.M. Zador, *Non-Gaussian Membrane Potential Dynamics Imply Sparse, Synchronous Activity in Auditory Cortex.* J. Neurosci., 2006. **26**(47): p. 12206-12218.

Chapter 3: Orchestrated excitatory and inhibitory plasticity enables stable persistent activity in a large spiking model

Abstract

The synaptic connectivity regime in which strong recurrent excitation is stabilized by inhibition enables neural circuits to sustain persistent activity states that are of fundamental computational utility. Yet the algorithmic rules that configure synaptic weights into this regime remain unknown. One excellent example of the self-organized emergence of stable persistent activity is the development of Up states in isolated cortical networks. Although experiments suggest that plasticity guides neurons to exhibit homeostatically-defined levels of activity, it has recently been shown that standard homeostatic rules alone are inherently poorly-suited to configure networks to support stable persistent activity states. Inspired by empirical observations of Up states, I model a large, sparsely-connected network of spiking units and compare the ability of different homeostatic-like plasticity rules to configure its synaptic weights in order to support stable persistent activity. I demonstrate that a family of cross-homeostatic plasticity rules can robustly configure an initially silent network's synaptic weights in order to generate Up states. Interestingly, I show that a variation on cross-homeostatic plasticity that uses local rather than global signals fails to properly configure networks toward stable Up states unless standard homeostatic plasticity operates alongside it. Results suggest that standard homeostatic and cross-homeostatic plasticity mechanisms can work together in a complementary manner to calibrate self-amplifying yet inhibition-stabilized circuits.

Chapter 3: Orchestrated excitatory and inhibitory plasticity enables stable persistent activity in a large spiking model

3.1 Introduction

It is well known that recurrent connectivity is the dominant circuit motif in brain areas such as cortex, which enables persistent activity states that are hypothesized to instantiate short-term memories among other computations[1, 2]. But it remains poorly understood how networks come to be able to exhibit stable persistent activity, particularly given the risk of uncontrollable runaway excitation (e.g. epileptic activity). What are the guiding principles that steer networks toward a regime capable of generating controlled persistent activity, without having their weights carefully pre-defined to do so? More concretely, what are the synaptic plasticity rules that allow inhibition-stabilized persistent activity to emerge in a self-organizing manner?

One of the best examples of self-sustained persistent activity, which is observed in cortical networks in many preparations *in vivo* and *in vitro*, comes from the bistable regime that switches between Down and Up states[3-6]. Up states refer to brief time periods (e.g. 500 ms - 2 s) in which nearly all neurons in the network fire spikes and become stably depolarized by recurrent volleys of synaptic communication. Experimental work has shown that cortical networks gradually transition from quiescence (i.e. near-zero firing rate) to the appearance of spontaneous Up states over the course of development both *in vivo*[7] and *in vitro*[8], during which time the Up states increase in frequency and duration. Interestingly, the spontaneous occurrence of Up states *in vitro* can be manipulated by chronically exciting pyramidal cells in the

culture for 24-48 hours[9, 10]. The latter result motivates the hypothesis that synaptic connections amongst excitatory and inhibitory neurons are guided by homeostasis – that is, the strength of synaptic connections are changed in order to achieve a target level of activity. Computational models have recently made progress in identifying the plasticity rules that could guide this process, by showing that certain homeostatic-like learning rules are capable of robustly modifying synaptic weights in order to support Up states [11].

However, there are several gaps in ongoing computational work. The orchestrated learning rules proposed by Soldado-Magraner and colleagues are composed as a sum of standard homeostatic and “cross-homeostatic” terms[11]. Standard homeostatic plasticity rules have long been well-defined[12] as an algorithm occurring within a single synapse that makes use of the sensed activity levels (e.g. firing rates) of only the presynaptic and postsynaptic neurons that form that synapse. In contrast, the cross-homeostatic term requires a neuron to maintain a sensed value of the activity of some set of neurons in the “opposite” population, i.e. an excitatory neuron would need to sense inhibitory firing rates, and vice versa. But does the postsynaptic neuron have access to a global signal representing the average activity of all cells in the opposite population, or does it only integrate the activity of the cells from which it receives input? Because the firing rate models that have been used so far were fully-connected, these two possibilities have been indistinguishable.

Moreover, it remains to be investigated whether the firing rate model’s results[11] depend on the assumptions and simplifications made relative to more biologically realistic spike-based models. In the current work, I hope to fill these gaps by

demonstrating that unsupervised excitatory and inhibitory learning rules can guide large, sparse networks of spiking units to exhibit persistent states with homeostatically-defined firing rates in the bistable regime. In grappling with the novel complexities of the model, I uncover nontrivial implementation and parameter effects, and explore their consequences.

3.2 Results

3.2.1 A large spiking network model of Up states

One of the best studied forms of emergent stable, self-sustained, and persistent activity states is the phenomenon of Up states [3-6], which occur in a bistable regime that alternates between Down and Up states. The Down state represents a state of network quiescence, with voltage near the resting membrane potential and firing rates near zero. In contrast, when an Up state is ignited nearly all neurons in the network become simultaneously active, propelled by massive volleys of synaptic communication that last around a second or longer. Despite the large-scale network activity during the Up state, average firing rates stay relatively low, and neurons are remarkably stably depolarized. How is this possible? I illustrate the answer to this question with a spiking model of Up states shown in **Fig. 3.1**.

In the model network shown in **Fig. 3.1**, 2000 leaky-adaptive integrate-and-fire units (1600 excitatory and 400 inhibitory, examples of spiking behavior shown in **Fig. 3.1A**) are connected on average to 25% of the other units with current-based synapses. Each unit thus receives on average 400 excitatory and 100 inhibitory synapses, with weights defined such that excitation dominates. When 100 excitatory units in the network receive an external excitatory input large enough to cause a single spike (a

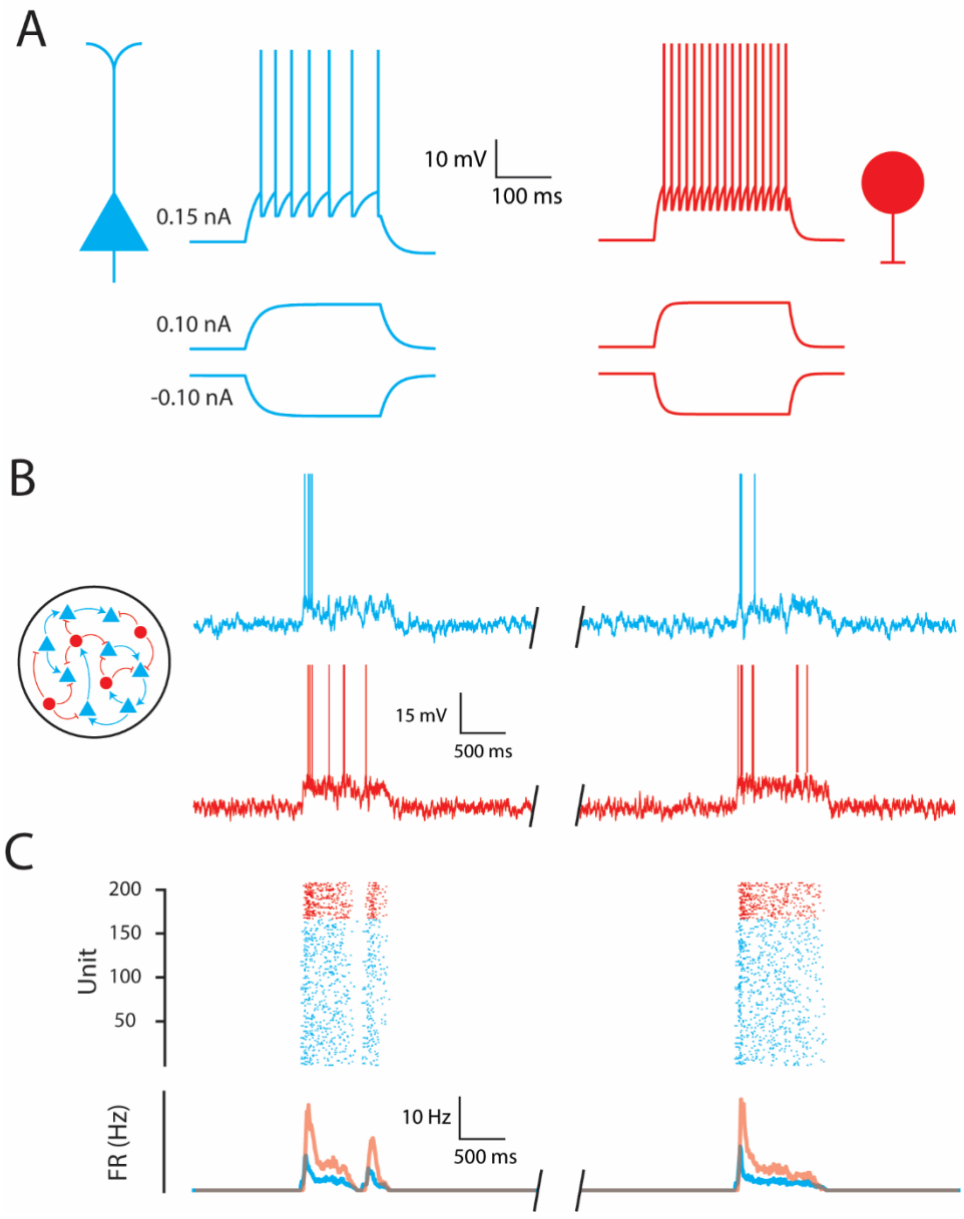

Figure 3.1 A large, sparse, spiking network model of Up states.
A. Units were leaky-adaptive integrate-and-fire units. Voltage traces display the response to 250 ms square pulses of injected current. Excitatory units (cyan) had a spike adaptation current, while inhibitory units (red) did not. **B.** When 100 excitatory units in the 2000-unit network (20% inhibitory) received external excitatory currents large enough to cause a spike, the network's recurrent connectivity (p_{conn} = 0.25) enabled Up states: stable and self-sustained bouts of recurrent activity lasting about one second. Exemplary voltage traces for individual excitatory and inhibitory units are shown for two Up states in a longer simulation. **C.** During simulated Up states, all units in the network become active, spiking at relatively low rates. The raster shows spiking activity for a random 200 unit subset of the network, while the lower traces show the PSTH averaged across all units in each population.

"kick")[13, 14], they recruit additional excitatory units, triggering an avalanche of activity (**Fig. 3.1B & C**). Yet the network does not explode (i.e. runaway excitation) because inhibitory units are also recruited, which provide negative currents that partially counteract excitatory currents. The net-positive current that each unit receives supports a stable firing rate, which in turns generates the amount of excitation and inhibition required to sustain itself. As the units continue to fire for several hundreds of milliseconds, a negative spike adaptation current builds in the excitatory units until it reaches a critical amount needed to destabilize the ongoing Up state[14], and the Up state terminates. These results demonstrate that the simulated network's behavior provides a sufficiently accurate model of Up states in the bistable regime when its synaptic weights are carefully pre-defined.

3.2.2 During an Up state, the model network exhibits the paradoxical effect

The key feature that allows a self-sustained state of persistent activity is the strong recurrent excitation that amplifies excitation produced by the "kick". In contrast, the key feature that enables stability is the recurrent inhibition that prevents the positive-feedback from "running away". When both features are in place, the network can be described as an inhibition-stabilized network, or ISN[15-17]. A signature behavior of ISNs is that externally-injected current applied to the inhibitory neurons causes the opposite of its expected effect, a phenomenon referred to as the paradoxical effect. As shown in **Fig. 3.2**, the spiking model of Up states exhibits the paradoxical effect. Here, I conducted a simulation in which I triggered an Up state that was made artificially more stable by disabling the spike adaptation current that is normally present in the excitatory units. During a 5 second Up state, a small positive current of variable size was injected

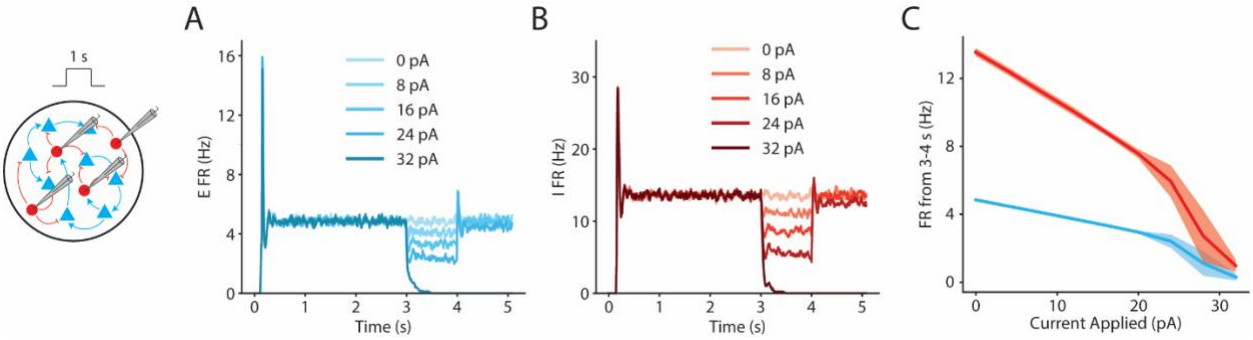

Figure 3.2 During an Up state, the model network exhibits the paradoxical effect. Depolarizing all inhibitory units during a simulated Up state decreased the mean firing rate (FR) of both excitatory (cyan) and inhibitory (red) populations. Up states were evoked in 5 second trials, and a square pulse of positive current was injected into the inhibitory units from 3 to 4 seconds. **A**. Post-stimulus time histograms (PSTHs) of the excitatory population average FR across 30 trials for each current amplitude. **B**. PSTHs of the inhibitory population average FR across trials. **C**. Mean FR from 3-4 s as a function of applied current. Increasingly depolarizing all inhibitory units decreases the FR for both populations.

into all of the inhibitory units in the network for 1 second (from 3-4 s). Thirty trials were conducted at each current value, and a post-stimulus time histogram (PSTH) was constructed by averaging the firing rate across trials and units for each population separately. Naively, one may expect that the injected positive current will depolarize inhibitory units, increase the inhibitory firing rate (FR), and that the increased FR in the inhibitory units will decrease the excitatory FR. However, in this network as the positive current injected into the inhibitory units is increased, the inhibitory FR *decreases*, and the excitatory FR decreases as well (**Fig. 3.2A & B**). A linear regression of the mean FR between 3 and 4 seconds under applied currents of 0, 8, 16, and 24 pA show that increasing the applied current predicted a decrease in the FR of both the excitatory and inhibitory populations (**Fig. 3.2C**; excitatory: β = -.10 Hz / pA; inhibitory: β = -.32 Hz / pA).

What explains this seemingly contradictory result? Perhaps most intuitively, the same self-amplifying recurrent excitation that endows the network with the ability to

exhibit persistence can also be self-deamplifying. More concretely, when positive current is injected into the inhibitory units, they briefly fire additional spikes that decreases the excitatory FR. Suppression of excitatory units then causes a withdrawal of excitation from both inhibitory and excitatory units. As long as the recurrent excitation that is withdrawn by the change in excitatory FR is larger than the external positive current being injected into the inhibitory units, the result will be a net decrease in the inhibitory FR. And this will only be the case in a network with strong recurrent excitation that is held in check by inhibition, i.e. an ISN.

The paradoxical effect can also be understood in dynamical terms if the spiking model is simplified into a Wilson-Cowan model[18], and the steady-state FR change from any given pair of initial FR values are visualized as a vector field in a two-dimensional plane with the average excitatory FR as the x-axis and inhibitory FR as the y-axis (i.e. standard phase-plane analysis)[19]. The positive current injected into the inhibitory units warps the phase-plane such that the upper stable fixed point – the point at which neither excitatory nor inhibitory FRs should change – is pushed toward the origin (i.e. both the excitatory and inhibitory FR at the fixed point decrease). And if the injected current is large enough (here, 32 pA), the phase plane will be warped such that the excitatory and inhibitory nullclines no longer intersect at the upper stable fixed point at all, ending the Up state.

3.2.3 The cross-homeostatic family of learning rules configures networks to support self-sustained persistent activity

In cortical organotypic cultures network behavior is initially completely dominated by the Down state, but over the course of *ex vivo* development Up states begin to occur

spontaneously, increasing in frequency and duration from the second to fourth week *in vitro*[8, 9, 11]. What are the algorithmic rules that govern this process? Studies have shown that the emergence of Up states can be manipulated by chronically exciting cells in the culture during the developmental period[9, 10], which suggests that the strength of synaptic connections amongst excitatory and inhibitory neurons are guided by homeostatic rules, i.e. rules that are designed to achieve a target level of activity. Recent computational work has shown, however, that the standard family of homeostatic plasticity rules is inherently unable to guide synaptic weights to regimes that support Up states[11]. Specifically, they can only successfully guide modeled networks to a self-sustained regime in a narrow and biologically implausible parameter region where inhibitory-to-excitatory synaptic strengths are initialized to be relatively large, and weights onto inhibitory units are changed much, much more slowly (e.g. 100 times more slowly) than weights onto excitatory units. In contrast, the same work demonstrated that a novel family of “cross-homeostatic” learning rules is inherently stable and robust to weight initialization while allowing excitatory and inhibitory weights to be changed at the same rate[11]. These rules consistently guided modeled networks to exhibit Up states over the course of simulated development.

Although the previous work showed that the cross-homeostatic family of plasticity rules is stable analytically and demonstrated its robustness in a fully-connected firing rate models 100 units[11], it has yet to be implemented in a sparsely-connected spiking model. Here, I used a simulated training protocol similar to what was used in the previous firing rate model, but I applied it to the spiking network model. Specifically, I provided the external kick that allowed recurrent excitation to ignite an Up state and

recorded the network's response in a 1500 ms trial. After calculating each unit's FR within the network response, synaptic weights were modified according to the cross-homeostatic learning rule, and the process was repeated over many trials.

Fig. 3.3 demonstrates a training session using the cross-homeostatic family of learning rules from an initially unresponsive network (i.e. a network dominated by the Down state). After a trial, each of the one million synaptic connections among the 1600 excitatory and 400 inhibitory units is modified according to the cross-homeostatic plasticity equations shown in **Fig. 3.3A** based on the network response to external stimulation. Initially, only the directly stimulated excitatory units fire spikes, so the other units receive a narrow volley of excitatory postsynaptic potentials (EPSPs) that lasts for tens of milliseconds (**Fig. 3.3B**, trial 1). Because the average FR of excitatory and inhibitory units during the network's response are far below their target setpoints, functionally excitatory synaptic weights (J_{EE} and J_{II}) are strengthened, while functionally inhibitory synaptic weights (J_{IE} and J_{EI}) are weakened (**Fig. 3.3D**, trials 1 to 200). Eventually, excitation from directly stimulated units pushes beyond the spike threshold of unstimulated units, driving reverberatory volleys of excitation and inhibition (**Fig. 3.3B**, trial 240). At this point, the network response contains complex polysynaptic activity that lasts up to a few hundred milliseconds even if it does not yet constitute a true Up state.

For the next phase of the training session (trial 300 to 500), recurrent excitation has become strong enough to support rudimentary persistent states, but evoked network dynamics are unstable and sputter out within several hundred milliseconds as excitatory units fire at a rate above their homeostatic setpoints (**Fig. 3.3B**, trial 320).

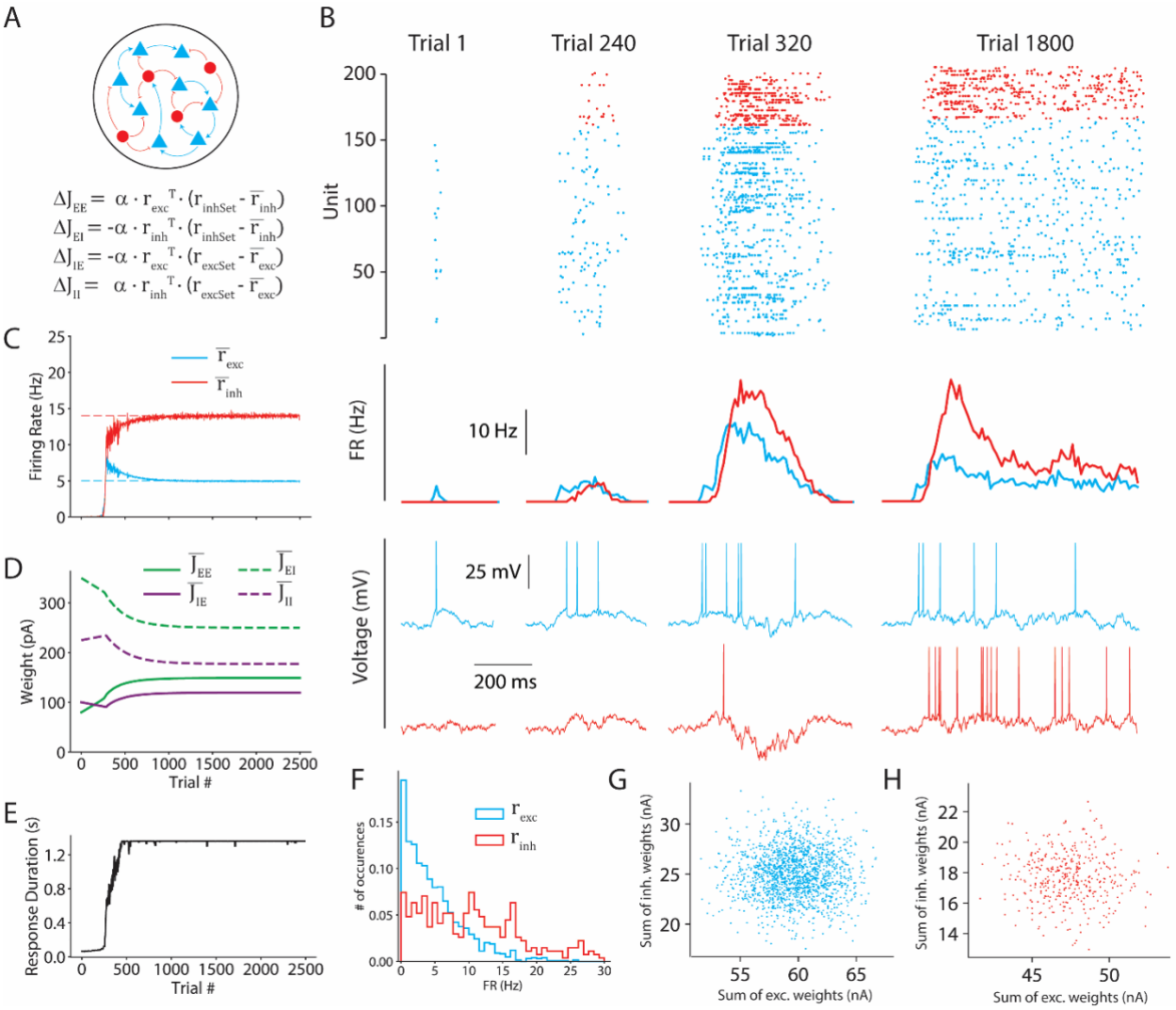

Figure 3.3 Cross-homeostatic learning rules configure synaptic weights to support Up states. **A**. The global implementation of the cross-homeostatic learning rule equations modifies synapses according to the presynaptic FR times the difference between the opposite population average FR and its homeostatic setpoint. **B.** Spike rasters, PSTHs, and voltage traces for excitatory (cyan) and inhibitory (red) units from exemplary trials over the course of training. **C**. Population average firing rates over the course of training. To aid in visualization, I show a moving average with a width of 5 trials. **D**. Mean weights for each synaptic class over the course of training. In this example the weights were initiated in an early developmental regime reflecting a silent network. **E**. Duration of the network spiking response over the course of training. To aid in visualization, I show a moving average with a width of 5 trials. **F**. Histograms of the unit FRs for the excitatory and inhibitory populations after the network has converged. **G**. Scatterplot of the sum of incoming excitatory synaptic weights versus the sum of incoming inhibitory synaptic weights for each excitatory unit. **H**. Same as G but for inhibitory units.

Accordingly, the vector describing the network's heading through weight space rotates: Although J_{EE} continues to strengthen and J_{EI} continues to weaken, J_{II} now begins to weaken while J_{IE} begins to strengthen. Unintuitively, this allows the network to calibrate the average excitatory FR downward toward its setpoint while continuing to boost the average inhibitory FR. As this process continues, the average excitatory FR settles into its setpoint (5 Hz) from above, while the average inhibitory FR arrives at its setpoint (14 Hz) from below. Over the course of training, network dynamics have become more stable and thus more persistent (**Fig. 3.3B**, trial 1800), and the response duration has progressively increased until it lasts for the full trial period considered here (**Fig. 3.3E**). Because the homeostatic error terms (e.g. $r_{inhSet} - \overrightarrow{r_{inh}}$) approach zero, weight changes become minimal, and the network remains in a final convergent state.

Importantly, neither the weights within a given class nor the individual unit FRs are homogeneous, as can be seen in the spike raster (**Fig. 3.3B**). Indeed, a histogram of unit FRs averaged across the final thousand trials (**Fig. 3.3F**) shows that they range widely with a coefficient of variation of about 0.8. This is because synaptic weights were initialized from a normal distribution, and each unit randomly receives a different amount of net current (i.e. balance of excitation and inhibition). Even though the cross-homeostatic learning rules are designed to modify weights so that the network reaches its setpoints, the rules importantly operate using the *average* FR of each population, i.e. the error term in each equation is essentially a single value that reflects the global error across all excitatory units or all inhibitory units. When the two populations have each reached their setpoints on average, weights will no longer change, so there is no motivation to balance excitation and inhibition onto an individual unit. Indeed, when the

final synaptic weight matrices are used to plot the sum of excitatory input onto each unit against the corresponding sum of inhibitory input, the relationship resembles a two-dimensional Gaussian distribution (**Fig. 3.3G & H**). In other words, the total excitation and inhibition that a given unit receives is not correlated (excitatory: β = 0.01, *p* = 0.80; inhibitory: β = -0.01, *p* = 0.72). Units in the upper left portion of the "cloud" that receive relatively large amounts of inhibition but little excitation will accordingly have a low FR, while those in the lower right portion that receive relatively little amounts of inhibition but large excitation will accordingly have a high FR.

3.2.4 A combined cross-homeostatic and homeostatic learning rule balances excitation and inhibition onto individual units

Because the error term in the implementation of the cross-homeostatic rule that I have shown thus far uses the global population mean firing rate (e.g. $r_{inhSet} - \overrightarrow{\overline{r_{inh}}}$), individual units are not driven to their setpoint FRs as in standard homeostatic learning rules. Although each population's *average* FR converges to the setpoint, there is a large variance in FR between units. For the same reason, if one were to assume that the cross-homeostatic rule is the only rule guiding unit FRs in a network, it would result in a peculiar experimental prediction: manipulating a small number of units in a network would not directly produce plasticity to help those units return to their own target setpoints (i.e. by modifying their own incoming synaptic weights) but rather would produce plasticity to help their postsynaptic partners reach their setpoints. This prediction ostensibly conflicts with experimental results showing, for example, that manipulating the FR of subsets of excitatory neurons in visual cortex causes the strength of inhibition onto the manipulated units to change in the same direction[20].

One potential resolution to this conflict is to assume that classical homeostatic forces are also in place, i.e. a learning rule formed by the sum of both cross-homeostatic and homeostatic terms (**Fig. 3.4A**). In fact, such a combined rule can be derived analytically from the loss function formed by the sum of excitatory and inhibitory unit error with respect to their setpoints [11].

Fig. 3.4 demonstrates the convergence process for what I will refer to as the "two-term" or "combined" family of plasticity rules, because it is formed as the sum of cross-homeostatic and homeostatic terms. Exact trajectories for the FR and weights differ from the purely cross-homeostatic case, with excitatory weights J_{EE} and J_{IE} increasing and inhibitory weights J_{EI} and J_{II} decreasing over the first several hundred trials in order to boost both the excitatory and inhibitory FRs (**Fig. 3.4B & C**). Once the excitatory FR has reached its setpoint, the inhibitory FR overshoots and all four weight classes reverse direction to optimize the inhibitory FR while maintaining the excitatory FR at its setpoint, eventually reaching a convergent and stable response of maximum duration (**Fig. 3.4D**). Convergence in population average FRs, average weights, and response duration superficially resembles the purely cross-homeostatic case, but there are some important differences. A histogram of the unit FRs averaged across the last thousand trials in the two-term training process shows that the variance in FR is drastically reduced compared to the purely cross-homeostatic case (**Fig. 3.4E versus 3.3G**), with a coefficient of variation of about 0.3 rather than 0.8. This difference in variance is statistically significant for both excitatory and inhibitory populations (excitatory: $F_{1599, 1599} = 4.8 \times 10^{-6}$, $p < .001$; inhibitory: $F_{399, 399} = 1.2 \times 10^{-6}$, $p < .001$). Plotting the sum of excitatory input onto each unit against the corresponding sum of

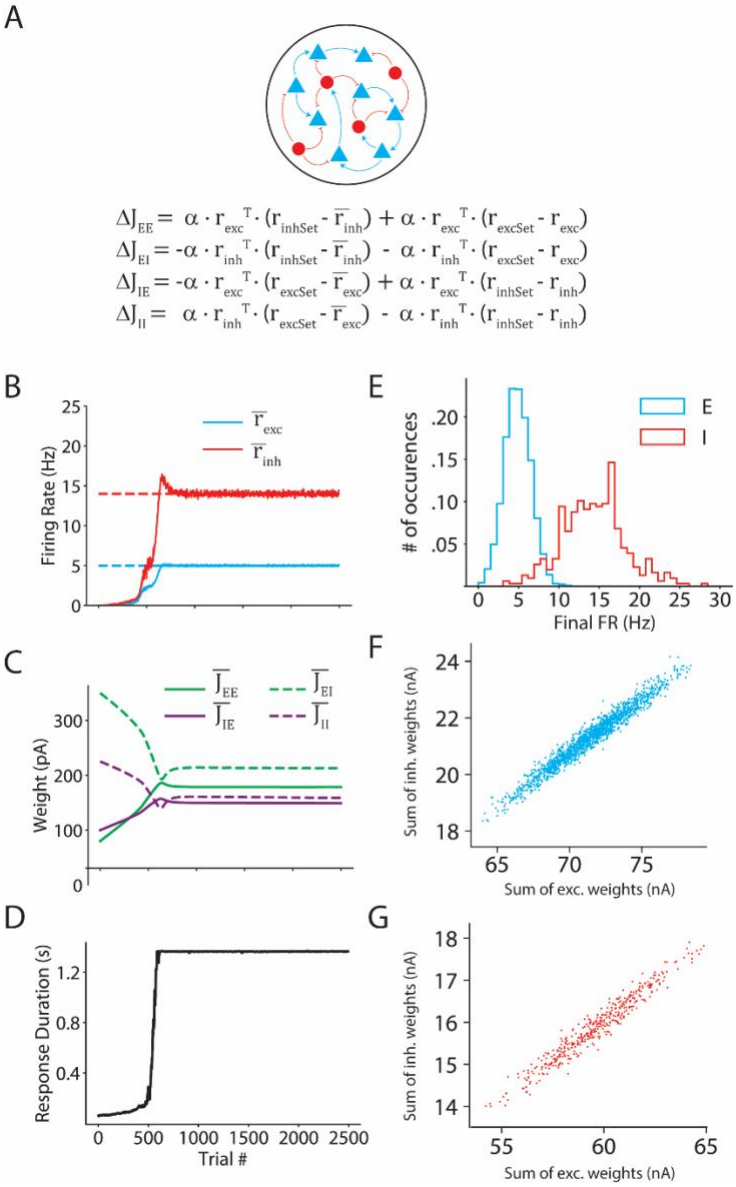

Figure 3.4 Combined cross-homeostatic and homeostatic learning rules balance excitation and inhibition onto individual units, decreasing cross-unit FR variance.
A. The homeostatic terms (second term in each equation) modify synapses according to the presynaptic FR times the difference between the postsynaptic FR and its homeostatic setpoint. **B**. Population average firing rates over the course of training. To aid in visualization, I show a moving average with a width of 5 trials. **C**. Mean weights for each synaptic class over the course of training. **D**. Duration of the network spiking response over the course of training. To aid in visualization, I show a moving average with a width of 5 trials. **E**. Histograms of the unit FRs for the excitatory and inhibitory populations after the network has converged. **F**. Scatterplot of the total incoming excitatory synaptic weights versus the total of incoming inhibitory synaptic weights for each excitatory unit. **G**. Same as F but for inhibitory units.

inhibitory input in the final weight matrices explains why this is the case. Unlike the purely cross-homeostatic case, the addition of the homeostatic term has caused excitatory and inhibitory input to individual units to become roughly balanced so that they resemble a line rather than a cloud (**Fig. 3.4F & G**). In these final weights, the inhibition that a given unit receives is correlated with the excitation that it receives (excitatory: $\beta = 0.38$, $p < 0.001$; inhibitory: $\beta = 0.37$, $p < 0.001$) In other words, when an individual unit receives a large amount of excitation, it also receives a large amount of inhibition, and when an individual unit receives a small amount of excitation, it also receives a small amount of inhibition. As a result, the net excitation that each unit receives is similar across units, and this is reflected in the tightened distributions of FRs.

The homeostatic rule balances excitation and inhibition onto individual units because, unlike the cross-homeostatic rule, it proposes a distinct weight change for each synapse that is intended to push the postsynaptic unit's individual FR to its setpoint. Although useful for balancing excitation and inhibition, the homeostatic rule alone fails to push networks toward weights that support stable and self-sustained activity states[11]. Yet when combined with the cross-homeostasis, the two rules can apparently work together without preventing the other from achieving its goal.

3.2.5 Local implementation of the cross-homeostatic rule fails unless counterbalanced by standard homeostasis

One potentially implausible property of the cross-homeostatic rule as I have implemented it thus far is that it assumes each neuron has access to a global signal representing the average activity of the opposite ("crossed") population. Although such a global signal may be possible, it may be more biologically plausible to assume that

any given neuron has access only to the activity of its presynaptic partners in the opposite population. This local variation on the cross-homeostatic rules can be implemented by simply replacing the variable that represents the average of the opposite population's FR (e.g. $\overrightarrow{r_{inh}}$) with a vector in which each element represents the average FR *amongst that unit's presynaptic partners* in the opposite population. In this case, proposed weight changes are distinct for each synapse depending on the hypoactivity or hyperactivity of its presynaptic partners in the opposite population. The left-hand column of panels in **Fig. 3.5** demonstrates an example training procedure using a local implementation of the cross-homeostatic rules. Interestingly, although the first several hundred trials resemble the global cross-homeostatic case, average FRs quickly become unstable and vary wildly from trial to trial (**Fig. 3.5B & C**). The average inhibitory FR overshoots its setpoint as the average excitatory FR undershoots its setpoint, and the average weights continue to diverge rather than reaching final steady values. If the training procedure is allowed to continue, the FRs and weights will never converge and instead continue to diverge, becoming more chaotic as the weights approach minimal and maximal values (not shown).

Although the exact cause for the failure of the local cross-homeostatic rule is not fully clear, two key observations help provide an intuition: First, examination of the individual unit FRs over the course of training shows that they vary chaotically; some units exhibit very large FRs for hundreds of trials only to eventually decrease back down to low FRs while other units do just the opposite (not shown). This is not the case in the global implementation, where although FR variance is high, units roughly maintain their FR over the course of training. Second, the local and global implementations affect FR

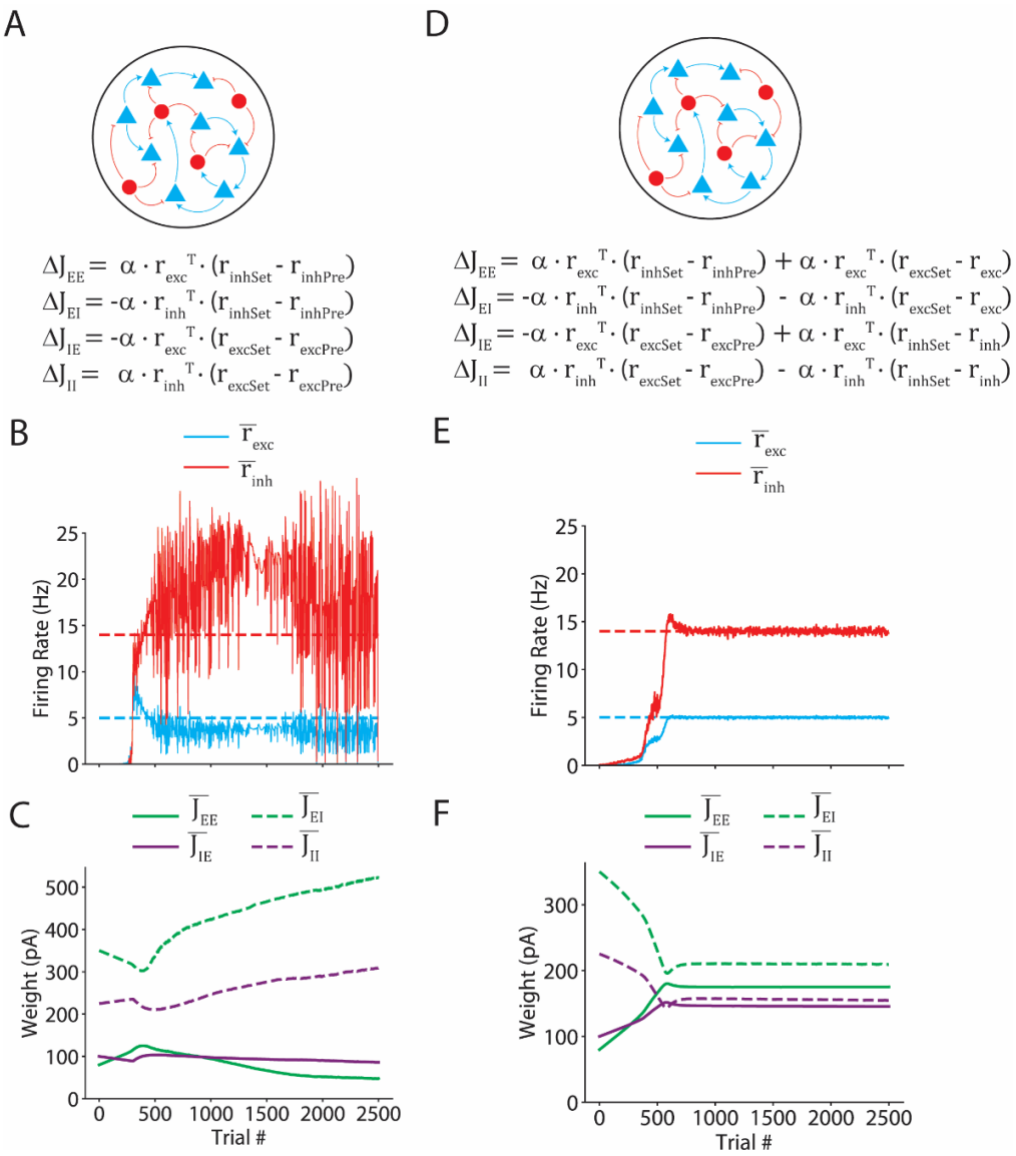

Figure 3.5 A local implementation of the cross-homeostatic learning rule fails unless the homeostatic rule operates alongside it.
A. The local implementation of the cross-homeostatic learning rule modifies synapses according to the presynaptic FR times the difference between the average FR among the presynaptic partners in the opposite population and their homeostatic setpoint. **B**. Population average firing rates over the course of training for the local cross-homeostatic rule alone. To aid in visualization, I show a moving average with a width of 5 trials. **C**. Mean weights for each synaptic class over the course of training for the local cross-homeostatic rule alone. **D**. The homeostatic terms (second term in each equation) modify synapses according to the presynaptic FR times the difference between the postsynaptic FR and its homeostatic setpoint. **E**. Population average firing rates over the course of training for the two-term rule with local cross-homeostasis. To aid in visualization, I show a moving average with a width of 5 trials. **F**. Mean weights for each synaptic class over the course of training for the two-term rule with local cross-homeostasis.

variance very differently over the course of training. Whereas the global implementation of the cross-homeostatic rule drives FR variance to asymptotically stable levels as the network successfully converges, the local implementation sharply increases FR variance to a higher value in both populations, and the increase in FR variance tracks the increasingly chaotic changes in average FR as the network fails to converge (not shown). Thus it seems that the way in which the local implementation proposes distinct changes to each synapse increases the cross-unit FR variance, and this problem becomes more and more exacerbated over the course of training.

As the right-hand column of panels in **Fig. 3.5** demonstrates, adding the classical homeostatic term to the local implementation of the cross-homeostatic rule allows it to successfully converge. Average FRs and weights follow a trajectory that strongly resembles the two-term rule with the global implementation of the cross-homeostatic rule (**Fig. 3.5E & F**). At the end of training, FR variance in the network that used the two-term rule with local cross-homeostasis is not statistically different from the network that used the two-term rule with global cross-homeostasis (excitatory: $F_{1599, 1599} = 1.05$, $p = .34$; inhibitory: $F_{399, 399} = 1.13$, $p = .24$). Thus, as in the global implementation, when the standard homeostatic rules are in place, they tend to decrease the cross-unit FR variance, which is apparently enough to counteract the local cross-homeostatic rule's tendency to uncontrollably increase cross-unit FR variance. Notably, combining the local cross-homeostatic and standard homeostatic rules yielded plasticity dynamics that guided the network toward stable self-sustained responses, despite each component rule failing to do so on its own.

3.2.6 Comparing convergence and stability across rule variations

In the current analysis, I have considered four distinct families of learning rules that result from the two possible implementations of the cross-homeostatic rule (global or local) and whether or not the homeostatic term was included (cross-homeostatic only or two-term). Although I illustrated their convergence in example training sessions that resemble developmental conditions (i.e. beginning from zero firing rate in both populations), it is also important to demonstrate the generality of these results across different weight initializations. In **Fig. 3.6**, I quantify the convergence and stability of the four different rules across 6000-trial training sessions beginning from nine different weight initializations.

To illustrate the nine weight initializations, I first show the trajectories taken by the two-term global family of rules through weight space. Weight initializations were chosen to capture a diversity of weight magnitudes (i.e. low-gain and high-gain) that started from either developmental conditions (zero FR) or nonzero FRs that were not at the setpoints. Despite starting from vastly different initial average weights, networks that used the rule families other than the local cross-homeostatic rule alone converged for all nine weight initializations. Ultimately, the networks' average weights arrived at different positions on a line reflecting the balance of excitation and inhibition that allowed them to exhibit sustained activity at the target setpoints (**Fig. 3.6A & B**).

In order to compare the ability of the different rules to guide network activity to the homeostatic setpoints and remain stable once convergence had taken place, I quantified the mean-squared error (MSE) of the population FR and the unit FRs with respect to their setpoints averaged over the last 2000 trials. First, I examined the

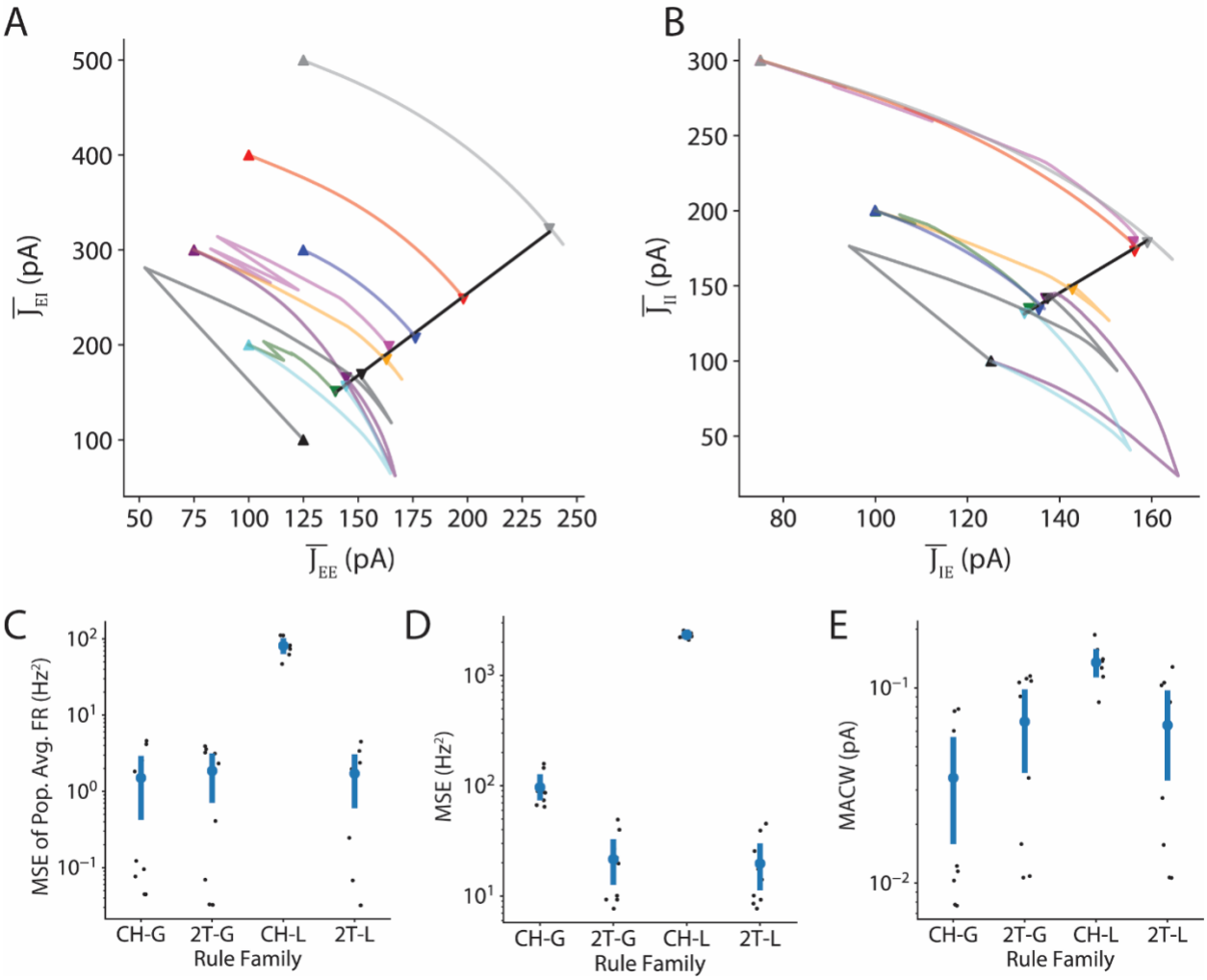

Figure 3.6 A comparison of convergence and stability across the four rule variations using different initial weight conditions at low and high gains.
A. A visualization of the initial mean weights onto the excitatory units (upward-pointing triangles) and the path of convergence taken by the 2T-global rule toward its final convergent weights (downward-pointing triangles) for the nine different initial conditions. **B.** Same visualization as in A, but for the mean weights onto the inhibitory units. **C**. The mean squared error (MSE) of the population average FRs with respect to their homeostatic setpoints over the final 2000 trials of training (i.e. after convergence) as a function of the rule family that was used. **D**. The MSE of individual unit FRs with respect to their homeostatic setpoints over the final 2000 trials of training as a function of the rule family that was used. **E**. The mean absolute change in weights (MACW) over the final 2000 trials of training as a function of the rule family that was used. For panels C, D, and E, all individual data points are shown, and error bars represent the 95% confidence interval. CH = cross-homeostatic, 2T = two-term, which refers to the combined homeostatic and cross-homeostatic rules, G = global cross-homeostasis, L = local cross-homeostasis.

population average FR with respect to its setpoint (**Fig. 3.6C**). I found that only the local cross-homeostatic rule failed to achieve an acceptable level of convergence with the population average FR. The mean terminal squared error for the population average FRs was 81.2 Hz^2 for the local cross-homeostatic rule, while it was about 1.7 Hz^2 for the other rules (post hoc t values all < -15, all p's < .001). There were no differences amongst the other three rules (all $p > 0.9$).

Next, I quantified the MSE of the individual unit FRs with respect to their setpoints (**Fig. 3.6D**). Since even for the two-term rules FR variance was considerable, this value was much larger in general. Unsurprisingly, I again found that the local cross-homeostatic rule performed the worst with a value of 2335 Hz^2 (post hoc t's all < -58, all p's < .001). In contrast, both two-term rules (with local and global cross-homeostasis) performed the best and equally well with values of 19.7 Hz^2 and 21.6 Hz^2 respectively, while the global cross-homeostatic rule's performance was intermediate with a value of 96.9 Hz^2. Post hoc tests from a one-way ANOVA that included the three successful rules showed that the two-term rules had significantly lower MSE of individual unit FRs (both t's < -7, p's < .001). This is because without the homeostatic term the global cross-homeostatic family of rules has no mechanism for driving individual units to their setpoints and instead can only optimize the population average FR toward the setpoint.

While the previous MSE measures were useful in quantifying the robustness of convergence across different weight initializations, I also wanted to quantify the stability of the rules once the target setpoints had been reached. To do so, I created a measure that captures how much weights changed on each trial, which I termed the mean absolute change in weights (MACW). I then quantified the average MACW across the

final 2000 trials. As shown in **Fig. 3.6E**, MACW was largest (mean value of 0.135 pA) for the local cross-homeostatic family of rules because it did not allow convergence and instead continued to make chaotic, unstable weight changes from trial to trial. MACW was significantly lower for the other three rules compared to the local cross-homeostatic rule (all *t*'s < -3.6, all *p*'s <.005). In contrast, the global cross-homeostatic learning rule had the lowest MACW with a mean value of 0.035 pA, and both two-term rules (with local and global cross-homeostasis) had intermediate and comparable mean values for the MACW of 0.064 and 0.067 pA. However, the differences in MACW amongst the three successful rules were not statistically significant (all $|t| < 1.7$, all *p*'s > 0.25). Although differences did not reach statistical significance, the global cross-homeostatic rule may be slightly more stable because once the population average FR reaches the setpoint, the error term approaches zero and the rule will not propose any more weight changes. On the other hand, the two-term rules could be considered slightly "less stable" than the global cross-homeostatic rule alone in the following way: from trial to trial, variations in random noise cause some units to fire more or less than their setpoints, and in response, the homeostatic term will attempt to optimize their FRs and continue to alter the weight matrices accordingly. Yet despite the ongoing attempt to optimize individual unit FRs, it is clear that the network's convergent behavior (i.e. exhibiting stable Up states with population average FRs at the setpoints) is maintained, and thus the combined rules can be said to be stable.

3.3 Discussion

Persistent activity states are among the most useful mechanisms for neural computation, because preserving information about a triggering event after it has

passed enables entire classes of cognitive functions. Since the identification of persistent activity states as a strong candidate mechanism underlying working memory[1], decades of experimental and computational work have explored its short-term mnemonic function and extended its role into arenas including timing, attention, and memory consolidation during sleep[21-23]. Although self-amplifying yet inhibition-stabilized regimes have been recognized as being critical for stable persistent activity, the algorithmic rules that can calibrate neural circuits into this regime remain poorly understood. In the current work I demonstrate that orchestrated plasticity rules based on homeostatic and cross-homeostatic principles allow large, sparsely-connected networks of spiking units to exhibit stable persistent activity states that resemble the well-studied phenomenon of Up states.

3.3.1 The interplay of cross-homeostatic and standard homeostatic forces

While useful for pushing each neuron toward its setpoint FR, standard homeostasis cannot cope with ISNs operating in the paradoxical regime[11]. Consider a hypothetical network that is pre-configured to support stable persistent activity in which all inhibitory units are firing just beneath their setpoint FR. Standard homeostatic plasticity would naively increase excitatory weights onto inhibitory units, but this strategy would counter-productively *decrease* inhibitory FR due to the paradoxical effect. This example illustrates one manifestation of how standard homeostasis can fail to configure networks to support stable persistent activity at the setpoint FRs [11]. Standard homeostasis can technically be stable when inhibitory-to-excitatory synaptic strengths are initialized to be relatively large, and inhibitory weights are changed much more slowly (e.g. 100 times more slowly) than excitatory weights. But experimental work

suggests that inhibitory neurons undergo homeostatic plasticity at the same rate as excitatory neurons[24-26], leading to the conclusion that standard homeostatic plasticity alone cannot account for the emergence of persistent stable activity. In contrast, as I demonstrate in **Fig. 3.3**, global cross-homeostatic rules can configure networks to exhibit stable Up states *by itself*. However, when interfacing the model with biological reality, the idea that global cross-homeostatic plasticity is the only learning rule at work in real networks faces at least two biological challenges.

First, if cross-homeostatic plasticity were operating alone it would lead to the prediction that manipulating the activity of certain subsets of units (e.g. some excitatory neurons) would not directly produce plasticity to help those units return to their FR setpoints. Instead of modifying their own incoming synapses, cross-homeostatic plasticity would instead modify synapses that target the opposite population, even if that population was already at its setpoint FR. Although such plasticity could conceivably modify FRs that would then cause additional plasticity to indirectly bring the manipulated units back to their setpoints, this prediction appears to conflict with previous experimental findings. For example, Xue and colleagues showed that after overexpressing a potassium channel in and consequently reducing the FR of a small subset of pyramidal cells in primary visual cortex, incoming inhibition onto the manipulated units was selectively reduced[20], which would be consistent with standard homeostatic plasticity.

Second, the global cross-homeostatic rule would require each unit in each population to have access to a signal representing the average FR in the opposite population. A simple modification to address this concern is to assume that each neuron

instead uses a local sensed value for the FR, e.g. by averaging the activity of its presynaptic partners in the opposite population. However, the current work predicts that using only local cross-homeostatic plasticity in sparse networks would chaotically destabilize plasticity dynamics throughout the population and prevent units from stably achieving their setpoints (**Fig. 3.5**, left column), which is also clearly inconsistent with experimental observations.

As the current work demonstrates, a combined rule formed by the sum of cross-homeostatic and standard homeostatic terms addresses both above concerns. Remarkably, such a combined rule can be derived analytically from the loss function formed by the sum of terms representing the squared error of excitatory and inhibitory populations with respect to their homeostatic setpoints[11]. Despite their differences, cross-homeostatic and homeostatic forces appear to work together in a complementary manner without interfering with each other. More specifically, the tendency of standard homeostatic plasticity to push each individual unit toward its setpoint by balancing its incoming excitation and inhibition – and thus decreasing cross-unit FR variance – counteracted the tendency of local cross-homeostasis to increase cross-unit variance. As a result, a combined rule with standard homeostasis and local cross-homeostasis successfully configured networks toward stable persistent activity at the setpoints (**Fig. 3.5**, right column), despite both component rules failing to do so on their own.

3.3.2 Cross-homeostasis: local or global?

As noted previously, the global cross-homeostatic rule would require each unit in either population to have access to a signal representing the average FR in the opposite population. Furthermore, two distinct global signals would be needed (one for each

population), and the value of the signal would need to be roughly homogeneous across the spatial extent of the population. However, the idea that neural networks use a global signal to regulate homeostasis may not be so far-fetched. For example, experimental results in hippocampal cultures showed that chronically reducing FRs through pharmacological means caused the network to adjust its average FR back to baseline, but individual neurons were not guaranteed to return to their baseline FRs[27]. A global signal could theoretically be implemented by a non-synaptic signaling molecule, e.g. a gaseous signaling molecule such as nitric oxide or a paracrine messenger[28]. In the absence of a global signaling mechanism, cross-homeostatic plasticity could also use purely local signals if a neuron can maintain a representation of the average FR across its presynaptic partners in the opposite population. One way to implement this biologically would be with metabotropic receptors, e.g. mGlu receptors on an inhibitory interneuron or $GABA_B$ receptors on a pyramidal cell. Slow activation of second messenger pathways could allow neurons to integrate synaptic activation from the opposite population, which could be compared to an internal, genetically-defined level of expected activation. And because homeostatic plasticity appears to primarily regulate ionotropic receptors, metabotropic receptor activation would be decoupled from ongoing plasticity.

Previous computational modeling work demonstrated that the cross-homeostatic family of plasticity rules can configure synaptic weights to support persistent stable activity in a firing rate model with 100 units[11]. However, because the network was fully-connected (each unit connected to all other units), the model was unable to distinguish between local and global implementations of the cross-homeostatic rule. In

the current work I used a sparsely-connected network (probability of connection = 0.25) and found, surprisingly, that when units sensed the FR of the opposite population locally via their synaptic inputs, the cross-homeostatic rule failed unless counter-balanced by standard homeostasis. Assuming that standard homeostatic plasticity also operates in parallel, the current results are thus consistent with either a global or local implementation of cross-homeostatic plasticity.

3.3.3 Future directions

A primary contribution of the current work was to show that conclusions made on the basis of simpler models (e.g. firing rate models) were upheld in a more biologically realistic model. Indeed, in computational models it is often difficult to determine the level of biological realism necessary to reach useful conclusions. For example, the sparse connectivity employed in my model was necessary to distinguish between local and global forms of cross-homeostasis. Nevertheless, there are some aspects of the current model that could be modified to be even more biologically realistic, which I would expect to be informative.

First, my model used only one type of inhibitory unit that was meant to resemble Parvalbumin interneurons and did not employ short-term synaptic plasticity (STP), a phenomenon in which a synapse's strength changes on short timescales based on its usage [29]. Experimental work has characterized several genetically distinct types of inhibitory interneurons with differing excitability and STP at incoming and outgoing synapses[30]. Recent computational work has shown that incorporating multiple inhibitory unit types[31] and endowing them with experimentally-observed STP profiles[32] impacts model behavior. Perhaps most relevant to the current work is the

observed predominance of weak short-term depression at excitatory-to-excitatory synapses[30], a force I expect to work alongside spike adaptation to place limitations on the duration of persistent states. Yet excitatory-to-excitatory STP has also been observed to be remarkably diverse and layer-dependent[33], which opens up the possibility that short-term facilitation could actually reinforce persistence in some parts of the circuit.

Second, the current work does not address the conundrum of how Up states (and other forms of stable persistent activity) are initiated. Spiking models of Up states generally throw this problem into stark relief, because one must decide how to model a stochastic source that can bridge the large (~15 mV) gap between resting membrane voltage and spike threshold[14]. If large amplitudes of membrane voltage noise are used, then voltage distributions are not bimodal as is observed experimentally, leading to the conclusion that large, synchronous events ("kicks") initiate Up states[13, 14]. But this leaves behind the mystery of what causes the "kicks." For *in vitro* cortical networks, possible candidate mechanisms including intrinsically-bursting pyramidal cells[34] or astrocytic calcium transients[35, 36].

3.3.4 Conclusion

Determining the learning rules that govern neural connectivity is a core goal in neuroscience because learning rules establish unifying principles that span molecular, cellular, systems, and computational levels of analysis. Here we reinforce and build on the theory of cross-homeostatic learning rules[11], showing that they can guide networks to support persistent activity states using local signals in a sparsely-connected

spiking network model, but only when operating alongside standard homeostatic learning rules.

3.4 Methods

3.4.1 Units

The units in the model were simulated as leaky integrate-and-fire neurons with a spike adaptation current. The membrane potential of each unit was represented as:

$$C_m \frac{dV(t)}{dt} = g_L(E_L - V(t)) + I_{syn}(t) - I_{adapt}(t) + \sigma\sqrt{\tau_m}\eta(t)$$

$$\frac{dI_{adapt}(t)}{dt} = \frac{-I_{adapt}(t)}{\tau_{adapt}}$$

The noise term $\sigma\sqrt{\tau_m}\eta(t)$ represents an Ornstein-Uhlenbeck process with zero mean, standard deviation σ, and a time constant equal to the membrane time constant $\tau_m = C_m/g_L$. When $V(t) \geq V_{thresh}$, the unit emitted a spike, its voltage was reset to V_{reset}, and its adaptation current I_{adapt} was incremented by β/τ_{adapt}. After spiking, the unit entered an absolute refractory period $\tau_{refractory}$. During a unit's absolute refractory period, it could not emit spikes, and the differential equation representing its voltage was not updated (i.e. its voltage was "paused" at V_{reset} until the absolute refractory period ended). Default values for unit parameters can be found in **Table 3.1**.

3.4.2 Synapses

At the core of my model were the interactions within and between the N_{exc} excitatory (E) units and N_{inh} inhibitory (I) units. Synapses were implemented as current-based, and the total synaptic current $I_{syn}(t)$ was summed across each unit's incoming synapses with distinct synaptic weights determined by the matrices J_{EE}, J_{IE}, J_{EI}, and J_{II}.

Thus the total synaptic current to the postsynaptic excitatory or inhibitory unit was given by each of the following two equations respectively:

$$I_{syn}(x,t) = \sum_{y=1}^{N_{exc}} J_{EE}(x,y)\, s_{syn}(x,y,t) + \sum_{y=1}^{N_{inh}} J_{EI}(x,y)\, s_{syn}(x,y,t)$$

$$I_{syn}(x,t) = \sum_{y=1}^{N_{exc}} J_{IE}(x,y)\, s_{syn}(x,y,t) + \sum_{y=1}^{N_{inh}} J_{II}(x,y)\, s_{syn}(x,y,t)$$

The kinetics of the synaptic currents were determined by the function $s_{syn}(x,y,t)$ for each presynaptic unit y and postsynaptic unit x. When a presynaptic spike occurred in unit y at time t^*, $s_{syn}(x,y,t)$ was incremented by an amount described by a delayed difference of exponentials equation[37]:

$$\Delta s_{syn}(x,y,t) = \frac{\tau_m}{\tau_d - \tau_r}\left[\exp\left(-\frac{t-\tau_l-t^*}{\tau_d}\right) - \exp\left(-\frac{t-\tau_l-t^*}{\tau_r}\right)\right]$$

where τ_m indicated the postsynaptic membrane time constant. Thus, the synaptic kinetics were determined by the synaptic delay τ_l, the synaptic rise time τ_r, and the synaptic decay time τ_d, which differed for excitatory and inhibitory synapses (see **Table 3.2**). Normalization constants were chosen so that varying synaptic time constants would not affect the time integral of the synaptic current. The synaptic delay τ_l was uniformly distributed between 0 and 1 ms (0 and 0.5 ms) across all excitatory (inhibitory) synapses. Default values for synaptic parameters can be found in **Table 3.2**.

3.4.3 Network

By default, networks consisted of N_{exc} E units (1600) and N_{inh} I units (400) with probability of connection $p_{conn} = 0.25$. Depending on p_{conn}, many of the values in the matrices J_{EE}, J_{IE}, J_{EI}, and J_{II} were zero indicating the absence of a connection. Each

Table 3.1 Unit parameters in Chapter 3.

Cell Parameter	Symbol	Value (E)	Value (I)	Unit
Resting potential	E_L	7.6	6.5	mV
Reset potential	V_{reset}	14	14	mV
Spike threshold	V_{thresh}	20	20	mV
Refractory period	$\tau_{refractory}$	5	2	ms
Membrane capacitance	C_m	200	100	pF
Leak conductance	g_L	10	10	nS
Membrane time constant	τ	20	10	ms
Adaptation strength	β	3	0	nA·ms
Adaptation time constant	τ_a	500	n/a	ms
Noise standard deviation	σ	2.5	2.5	mV

Model parameters defining intrinsic properties of excitatory (E) and inhibitory (I) units.

Table 3.2 Synaptic parameters in Chapter 3.

Synaptic Parameter	Symbol	Value (E)	Value (I)	Unit
Rise time	τ_r	8	1	ms
Fall time	τ_d	23	1	ms
Mean synaptic delay	τ_l	1	0.5	ms

Model parameters defining kinetics of excitatory (E) and inhibitory (I) synapses.

weight matrix was generated by first initializing a connectivity matrix A_{XY} of zeros, with N_x columns and N_y rows. After calculating the expected number of synapses that would be present based on p_{conn}, that many elements in A_{XY} were chosen without replacement to be 1s. Nonzero diagonal elements were not allowed in the case of A_{EE} or A_{II}. A_{XY} thus always had uniformly random connectivity with no autapses and a probability of connection exactly equal to p_{conn}. Weight matrices were then initialized by replacing the nonzero elements of A_{XY} with numbers drawn from normal distributions of a pre-defined mean and a coefficient of variation equal to 0.2. Network simulations were evaluated using forward Euler integration using a time step of 0.1 ms.

3.4.4 Procedure

Because the homeostatic learning rules that have been developed thus far are designed to optimize the firing rate (FR) during the Up state for the E and I units, I used a procedure in which Up states were triggered in short trials of 1.5 seconds. In each trial, a synaptic current large enough to cause a spike ($I_{syn} \Rightarrow I_{syn} + 0.98\, nA$) was injected into a number of E units equal to 5% of the total population (i.e. for a network with 2000 units, 100 E units). This constituted a “kick” that provided the possibility for recurrent excitation to ignite an Up state.

After each trial, I first calculated the post-stimulus time histogram (PSTH) of spiking across all units of each type as the average number of spikes per unit in 10 ms bins, yielding a population PSTH for the excitatory and inhibitory units over the course of the trial. Using the population PSTH of the I units, Up states were detected as contiguous periods of time in which the FR exceeded 0.2 Hz for at least 100 ms. Down states that were shorter than 100 ms were considered interruptions of an Up state and

were thus “deleted,” combining the surrounding Up states. In general, I found that this method of detecting Up states was robust, and many parameters worked well to detect Up states. Further, because I used relatively short 1500 ms trials with a single “kick” presented at 100 ms, it was extremely rare that more than one Up state would be elicited (i.e. a single Up state was either elicited or not).

If there was a single Up state, the FR during the Up state for each unit was calculated in order to be used in the learning rule. If there were no Up states, I instead calculated the average FR for each unit in the time period of nonzero FR, although in this case I considered the FR of the 100 directly stimulated excitatory units to be the average of all other excitatory units in order to avoid synaptic weight biases onto and from the input population. FRs for each unit in each trial contributed to a vector of moving averages with time constant τ_{FR} (here, $\tau_{FR} = 2$). The moving average was initialized at the exact FR elicited during the first trial. Using the moving averages, I constructed the row vectors $\vec{r}_{exc}$ and $\vec{r}_{inh}$ in which the elements represent the moving average values of each unit’s FR.

3.4.5 Learning rules

Standard homeostatic plasticity. The homeostatic learning rules considered here operate on the assumption that each type of unit (excitatory or inhibitory) has a fixed “set-point” of activity, which are scalars that I refer to here as r_{excSet} and r_{inhSet}. Based on experimental data[31] I used values of 5 and 14 Hz for the r_{excSet} and r_{inhSet} respectively. Weight changes are applied between trials on the basis of the comparison between the observed unit FRs during the Up state ($\vec{r}_{exc}$ and $\vec{r}_{inh}$) and the setpoints.

I first describe a family of learning rules that I refer to as the *standard homeostatic* family, in which a synapse's strength is modified based on the product between the presynaptic FR and the "error" of the postsynaptic unit's FR with respect to the target set-point:

$$\Delta J_{EE} = \alpha_1 \cdot \vec{r}_{exc}{}^{\mathrm{T}} \cdot (r_{excSet} - \vec{r}_{exc})$$

$$\Delta J_{EI} = -\alpha_1 \cdot \vec{r}_{inh}{}^{\mathrm{T}} \cdot (r_{excSet} - \vec{r}_{exc})$$

$$\Delta J_{IE} = \alpha_1 \cdot \vec{r}_{exc}{}^{\mathrm{T}} \cdot (r_{inhSet} - \vec{r}_{inh})$$

$$\Delta J_{II} = -\alpha_1 \cdot \vec{r}_{inh}{}^{\mathrm{T}} \cdot (r_{inhSet} - \vec{r}_{inh})$$

where α_1 is a learning rate constant, with units that convert the product of FRs into the same units as the weight matrix (in the current study, α_1 is defined by default to be $0.0025 \frac{pA}{Hz^2}$). Notice that the signs above have been chosen so that if the postsynaptic unit's FR is below its setpoint, excitation onto that unit is increased while inhibition onto that unit is decreased. In contrast, when the postsynaptic unit's FR is above its setpoint, excitation onto that unit is decreased while inhibition onto that unit is increased.

Cross-homeostatic plasticity. The *cross-homeostatic* family[11] rely on the product between the presynaptic FR and the average FR error of a set of units in the opposite or "cross" population:

$$\Delta J_{EE} = \alpha_2 \cdot \vec{r}_{exc}{}^{\mathrm{T}} \cdot (r_{inhSet} - \vec{r}_{inhCross})$$

$$\Delta J_{EI} = -\alpha_2 \cdot \vec{r}_{inh}{}^{\mathrm{T}} \cdot (r_{inhSet} - \vec{r}_{inhCross})$$

$$\Delta J_{IE} = -\alpha_2 \cdot \vec{r}_{exc}{}^{\mathrm{T}} \cdot (r_{excSet} - \vec{r}_{excCross})$$

$$\Delta J_{II} = \alpha_2 \cdot \vec{r}_{inh}{}^{\mathrm{T}} \cdot (r_{excSet} - \vec{r}_{excCross})$$

where α_2 is again a learning rate constant with the same units as α_1. To implement the error term ($\vec{r}_{popCross}$) I examined two distinct formulations. First, and most simply, cross-homeostatic learning could rely on a global signal that integrates and averages the activity of all units in the opposite population. In this case, $\vec{r}_{popCross}$ is a column vector of length equal to the number of units in the postsynaptic population, with all values set to the average of the opposite population:

$$\vec{r}_{inhCrossGlobal}(i) = mean(\vec{r}_{inh})\ for\ i = 1, 2, \dots, N_{exc}$$

$$\vec{r}_{excCrossGlobal}(i) = mean(\vec{r}_{exc})\ for\ i = 1, 2, \dots, N_{inh}$$

In the second formulation, cross-homeostatic plasticity relied upon a local signal, in which each postsynaptic unit only has access to the activity of its presynaptic partners in the opposite population. In this case, each element of $\vec{r}_{popCross}$ represents the average FR of the units in the opposite population that synapse onto that unit. For this implementation I multiply the unit FRs by the connectivity matrix A_{XY} and divide by the vector that results from summing its columns, which I refer to as $\vec{a}_{XY}$. Note that the $\oslash$ symbol refers to element-wise division.

$$\vec{r}_{inhCrossLocal} = A_{EI}\vec{r}_{inh} \oslash \vec{a}_{EI}$$

$$\vec{r}_{excCrossLocal} = A_{IE}\vec{r}_{exc} \oslash \vec{a}_{IE}$$

Two-term cross-homeostatic plasticity. Lastly, I consider a combined homeostatic and cross-homeostatic learning rule in which the two families defined above are summed:

$$\Delta J_{EE} = \alpha_1 \cdot \vec{r}_{exc}^{\,\mathrm{T}} \cdot (r_{excSet} - \vec{r}_{exc}) + \alpha_2 \cdot \vec{r}_{exc}^{\,\mathrm{T}} \cdot (r_{inhSet} - \vec{r}_{inhCross})$$

$$\Delta J_{EI} = -\alpha_1 \cdot \vec{r}_{inh}^{\,\mathrm{T}} \cdot (r_{excSet} - \vec{r}_{exc}) - \alpha_2 \cdot \vec{r}_{inh}^{\,\mathrm{T}} \cdot (r_{inhSet} - \vec{r}_{inhCross})$$

$$\Delta J_{IE} = \alpha_1 \cdot \vec{r}_{exc}^{\,\mathrm{T}} \cdot (r_{inhSet} - \vec{r}_{inh}) - \alpha_2 \cdot \vec{r}_{exc}^{\,\mathrm{T}} \cdot (r_{excSet} - \vec{r}_{excCross})$$

$$\Delta J_{II} = -\alpha_1 \cdot \vec{r}_{inh}^{\,\mathrm{T}} \cdot (r_{inhSet} - \vec{r}_{inh}) + \alpha_2 \cdot \vec{r}_{inh}^{\,\mathrm{T}} \cdot (r_{excSet} - \vec{r}_{excCross})$$

Finally, for all terms in all learning rules, whenever there appears a vector of presynaptic FRs ($\vec{r}_{exc}$, $\vec{r}_{inh}$), I imposed a minimum value such that each element was at least 1 Hz. The minimum constraint (1 Hz) was instituted because the weight changes are always a product of the presynaptic FR; thus, when the presynaptic FRs are near zero, the weight changes would be near zero and would take hundreds of times more trials to make significant changes, which was computationally impractical. Note that this minimum value constraint was not applied to the appearance of FR vectors when they occur in a homeostatic error term. Additionally, synaptic weights were constrained to stay within minimum and maximum weight values of 10 pA and 750 pA respectively, for all synapses.

3.4.6 Evaluations of learning rules: convergence and post-convergence stability

To evaluate the ability of the learning rules to push the network toward the setpoints, I first used a standard mean squared error (MSE) approach and averaged the squared error of the excitatory and inhibitory population average FRs at each trial:

$$MSE_{pop}(trial) = \frac{1}{2}(mean(\vec{r}_{exc}) - r_{excSet})^2 + \frac{1}{2}(mean(\vec{r}_{inh}) - r_{inhSet})^2$$

Next, to quantify error of individual units with respect to their setpoints, I created a quantifier of MSE across all individual excitatory and inhibitory units at each trial:

$$MSE_{units}(trial) = \frac{1}{N_{pop}} \sum_{i=1}^{N_{pop}} (\vec{r}_i - r_{popSet})^2$$

where $\vec{r}_{pop}$ represented the vector of FRs of that population (excitatory or inhibitory) and r_{popSet} represented the corresponding set-point.

Finally, because convergence to the setpoints does not necessarily ensure that the weights have “converged” (i.e. stopped changing) I additionally employed a measure of the mean absolute change in weights on a given trial:

$$MACW(trial) = mean(|\Delta J_{EE}|) + mean(|\Delta J_{EI}|) + mean(|\Delta J_{IE}|) + mean(|\Delta J_{II}|)$$

3.5 Chapter 3 References

1. Goldman-Rakic, P.S., *Cellular basis of working memory.* Neuron, 1995. **14**(3): p. 477-85.

2. Wang, X.J., *Synaptic reverberation underlying mnemonic persistent activity.* Trends Neurosci, 2001. **24**(8): p. 455-63.

3. Timofeev, I., F. Grenier, and M. Steriade, *Disfacilitation and active inhibition in the neocortex during the natural sleep-wake cycle: An intracellular study.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(4): p. 1924-1929.

4. Shu, Y., A. Hasenstaub, and D.A. McCormick, *Turning on and off recurrent balanced cortical activity.* Nature, 2003. **423**(6937): p. 288-293.

5. Maclean, J., et al., *Internal Dynamics Determine the Cortical Response to Thalamic Stimulation.* Neuron, 2006. **48**: p. 811-23.

6. Compte, A., et al., *Cellular and Network Mechanisms of Slow Oscillatory Activity (<1 Hz) and Wave Propagations in a Cortical Network Model.* Journal of Neurophysiology, 2003. **89**(5): p. 2707-2725.

7. Golshani, P., et al., *Internally Mediated Developmental Desynchronization of Neocortical Network Activity.* The Journal of Neuroscience, 2009. **29**(35): p. 10890-10899.

8. Johnson, H.A. and D.V. Buonomano, *Development and Plasticity of Spontaneous Activity and Up States in Cortical Organotypic Slices.* Journal of Neuroscience, 2007. **27**(22): p. 5915-5925.

9. Motanis, H. and D. Buonomano, *Delayed in vitro development of Up states but normal network plasticity in Fragile X circuits.* European Journal of Neuroscience, 2015. **42**(6): p. 2312-2321.

10. Goel, A. and D.V. Buonomano, *Chronic electrical stimulation homeostatically decreases spontaneous activity, but paradoxically increases evoked network activity.* J Neurophysiol, 2013. **109**(7): p. 1824-36.

11. Soldado-Magraner, S., et al., *Orchestrated Excitatory and Inhibitory Learning Rules Lead to the Unsupervised Emergence of Up-states and Balanced Network Dynamics.* bioRxiv, 2021: p. 2020.12.30.424888-2020.12.30.424888.

12. Turrigiano, G.G., et al., *Activity-dependent scaling of quantal amplitude in neocortical neurons.* Nature, 1998. **391**(6670): p. 892-896.

13. DeWeese, M.R. and A.M. Zador, *Non-gaussian membrane potential dynamics imply sparse, synchronous activity in auditory cortex.* Journal of Neuroscience, 2006. **26**(47): p. 12206-12218.

14. Jercog, D., et al., *UP-DOWN cortical dynamics reflect state transitions in a bistable network.* eLife, 2017. **6**: p. 1-33.

15. Tsodyks, M.V., et al., *Paradoxical Effects of External Modulation of Inhibitory Interneurons.* The Journal of Neuroscience, 1997. **17**(11): p. 4382-4388.

16. Ozeki, H., et al., *Inhibitory Stabilization of the Cortical Network Underlies Visual Surround Suppression.* Neuron, 2009. **62**(4): p. 578-592.

17. Sanzeni, A., et al., *Inhibition stabilization is a widespread property of cortical networks.* bioRxiv, 2019: p. 1-39.

18. Wilson, H.R. and J.D. Cowan, *Excitatory and inhibitory interactions in localized populations of model neurons.* Biophysical journal, 1972. **12**(1): p. 1-24.

19. Gerstner, W., et al., *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition*. 2014: Cambridge University Press.

20. Xue, M., B.V. Atallah, and M. Scanziani, *Equalizing excitation–inhibition ratios across visual cortical neurons.* Nature, 2014. **511**(7511): p. 596-600.

21. Major, G. and D. Tank, *Persistent neural activity: Prevalence and mechanisms.* Current Opinion in Neurobiology, 2004. **14**(6): p. 675-684.

22. Compte, A., *Computational and in vitro studies of persistent activity: Edging towards cellular and synaptic mechanisms of working memory.* Neuroscience, 2006. **139**(1): p. 135-151.

23. Zylberberg, J. and B.W. Strowbridge, *Mechanisms of Persistent Activity in Cortical Circuits: Possible Neural Substrates for Working Memory.* Annual Review of Neuroscience, 2017. **40**: p. 603-627.

24. Kuhlman, S.J., et al., *A disinhibitory microcircuit initiates critical-period plasticity in the visual cortex.* Nature, 2013. **501**(7468): p. 543-546.

25. Gainey, M.A., J.W. Aman, and D.E. Feldman, *Rapid disinhibition by adjustment of pv intrinsic excitability during whisker map plasticity in mouse s1.* Journal of Neuroscience, 2018. **38**(20): p. 4749-4761.

26. Ma, Z., et al., *Cortical Circuit Dynamics Are Homeostatically Tuned to Criticality In Vivo.* Neuron, 2019. **104**(4): p. 655-664.e4.

27. Slomowitz, E., et al., *Interplay between population firing stability and single neuron dynamics in hippocampal networks.* eLife, 2015. **2015**(4): p. 1-21.

28. Drew, P.J., *Vascular and neural basis of the BOLD signal.* Current Opinion in Neurobiology, 2019. **58**: p. 61-69.

29. Zucker, R.S. and W.G. Regehr, *Short-Term Synaptic Plasticity.* Annual Review of Physiology, 2002. **64**(1): p. 355-405.

30. Campagnola, L., et al., *Connectivity and Synaptic Physiology in the Mouse and Human Neocortex.* bioRxiv, 2021.

31. Romero-Sosa, J.L., H. Motanis, and D.V. Buonomano, *Differential excitability of PV and SST neurons results in distinct functional roles in inhibition stabilization of Up-states.* bioRxiv, 2020. **2**(310): p. 1-27.

32. Seay, M.J., et al., *Differential Short-Term Plasticity of PV and SST Neurons Accounts for Adaptation and Facilitation of Cortical Neurons to Auditory Tones.* The Journal of neuroscience : the official journal of the Society for Neuroscience, 2020. **40**(48): p. 9224-9235.

33. Lefort, S. and C.C.H.H. Petersen, *Layer-Dependent Short-Term Synaptic Plasticity between Excitatory Neurons in the C2 Barrel Column of Mouse Primary Somatosensory Cortex.* Cerebral Cortex, 2017. **27**(7): p. 3869-3878.

34. Neske, G.T., S.L. Patrick, and B.W. Connors, *Contributions of Diverse Excitatory and Inhibitory Neurons to Recurrent Network Activity in Cerebral Cortex.* Journal of Neuroscience, 2015. **35**(3): p. 1089-1105.

35. Poskanzer, K.E. and R. Yuste, *Astrocytic regulation of cortical UP states.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(45): p. 18453-18458.

36. Poskanzer, K.E. and R. Yuste, *Astrocytes regulate cortical state switching in vivo.* Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(19): p. E2675-E2684.

37. Brunel, N. and X.J. Wang, *What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance.* J Neurophysiol, 2003. **90**(1): p. 415-30.

Chapter 4: Testing the relationship between timing and working memory with two complementary tasks

Abstract

Working memory and timing have generally been considered to be separate mental functions with distinct neural implementations. Yet both require transiently storing information for future usage – retrospective information in the case of working memory and prospective information in the case of timing. To test the hypothesis that working memory and timing are linked, I designed two behavioral tasks that had the same stimulus structure but differed in terms of whether working memory or timing information was explicitly required to respond correctly. In both cases, participants learned about and used the other task-irrelevant component. The pattern of results is consistent with the employment of cue-specific neural sequence-like representations that multiplex working memory and timing information. This suggests that in some cases working memory is coded in a time-varying format because of the importance of predicting when it will be used.

Chapter 4: Testing the relationship between timing and working memory with two complementary tasks

4.1 Introduction

In daily life, humans engage with events that have predictable temporal structure. For example, the moment a traffic stoplight changes from green to yellow, one can predict about how long it will take for the light to subsequently change to red. Furthermore, our ability to prospectively anticipate events is also flexible, as the particular traffic intersection, the type of stoplight (e.g. yellow arrow), and the context (e.g. as a pedestrian rather than driver) could determine the relevant duration and/or the best behavioral response. It is clear that such a feat depends upon *timing*, which I use here to refer to the ability to track time to anticipate the onset of external events on the scale of seconds[1, 2]. Yet one can also make the argument that such an anticipatory act depends on *working memory* (WM), the ability to transiently maintain information in mind (e.g., the light changed) and later use it flexibly to support a behavior or decision [3, 4]. Indeed, the example illustrates how prospectively timing a predictable interval and retrospectively maintaining information about the initiating event are often intertwined. Without retrospective information, a timing prediction would lose its particular meaning and thus its utility for flexibly guiding behavior at a future moment.

A recently growing body of evidence suggests that WM and timing are interconnected computations. In nonhuman animals, studies have observed similar signatures of neural activity during the fixed delay periods in timing tasks such as temporally-cued reward[5, 6] and in WM tasks such as delayed nonmatch-to-sample[7]. In diverse areas including hippocampus, striatum, and cortex, neural activity patterns

during the delay period resemble *sequences*, in which single neurons fire sparsely at certain times, i.e. the peak firing rates throughout the neural population “tile” the delay period. Computational and analytical work suggests that sequential population activity provides an optimal representational framework for tracking time[6], akin to temporal basis functions used in temporal difference reinforcement learning[8]. There is also evidence that neural activity during the delay period of WM tasks is more generally time varying[9-12], including cue-specific neural sequences [7]. But the coding framework that is used to simultaneously track time and preserve cue identity – and how the computational strategy depends on the exact behavioral requirements – remains unclear.

If time-varying activity (e.g. a cue-specific sequence) is used to instantiate a fused representation of both WM and timing, then one would predict that requiring mnemonic information at an unexpected time (e.g. halfway through the sequence) would yield weaker performance than at the expected time. Indeed, behavioral studies in humans have demonstrated that while performing a WM task, participants make slower and less accurate memory judgments when probed at an unexpected time [13-15]. A related prediction would be that extracting timing information from a cue-specific neural sequence should result in systematic errors if that cue is typically associated with a different time interval. For example, if a cue associated with a short delay is instead followed by a long delay, then the neural state triggered by the cue may provide timing information that is incongruent with reality. To my knowledge, this prediction has not been tested.

In the current work, I developed two behavioral tasks that use the same stimulus structure but vary as to whether the WM or timing components are explicit (i.e. required to solve the task) or implicit (i.e. task-irrelevant, although potentially informative). My results demonstrate that participants learn task-irrelevant timing information during an explicit WM task and task-irrelevant WM information during an explicit timing task. I add to existing evidence that probing WM at an unexpected time impairs performance, and I provide novel evidence that probing internal timing at moments that conflict with the prediction of a timing-informative cue leads to systematic errors in time judgement.

4.2 Results

4.2.1 The differential delayed match-to-sample (dDMS) task

To test the relationship between timing and WM, I first developed a variant of a standard WM task, the delayed match-to-sample task. Delayed match-to-sample tasks involve three phases: cue, delay, and probe. First, the participant is presented with a cue stimulus, in the simplest case with only two alternatives (A or B). The cue is then removed, and participants must maintain the identity of the cue in mind during an empty delay period. Finally, once the delay period has ended, either of the two stimuli are presented as a probe, resulting in four possible cue-probe combinations (AA, AB, BA, BB). Once the probe has been presented, participants are required to respond differentially based on whether the cue and probe matched (AA or BB) or did not match (AB or BA). In most prior research, the duration of the delay period between the cue stimulus and probe stimulus is fixed or randomized across trials; however, the key variation in the current task was that cue identity predicted the duration of the delay period on most trials (**Fig. 4.1A**). Accordingly, I termed the task the differential delayed

match-to-sample (dDMS) task. For example, if the cue was A (AA or AB), the delay period was likely to last 1 second (80% of trials), while if the cue was B (BA or BB), the delay period was likely to last 2.2 seconds (80% of trials), and this mapping was counterbalanced across participants. Importantly, the duration of the delay period was irrelevant to the requirements and instructions of the task. I hypothesized that through experience with the task, participants would form an association between the cue identity and the delay period duration, allowing them to predict when the probe would appear based on the cue – or conceivably, to infer the content of WM based on when the probe appeared. To test whether participants learned the cue-delay association, the cue-delay contingency was reversed in a random 20% of the trials, allowing me to contrast behavioral performance between what I refer to as the Standard and Reverse trials.

Three behavioral measures were used to quantify WM performance in the task: median reaction time (RT), accuracy (proportion of trials with correct responses), and the inverse efficiency score (IES) – a measure that combines information about both RT and accuracy. IES, which is simply RT divided by accuracy, takes into account the possibility that individual participants may prioritize speed at the expense of accuracy or *vice versa*, also known as the speed-accuracy tradeoff[16, 17]. For all three measures, there was a main effect of Reversal in which behavioral performance was worse for the Reverse trials (RT: $F_{1, 26} = 7.4$, $p < .05$; accuracy: $F_{1, 26} = 13.4$, $p < .005$; IES: $F_{1, 26} = 9.1$, $p < .01$), indicating that violation of the cue-delay association impaired WM performance (**Fig. 4.1B-D**). In other words, participants were not only slower when WM was probed at the unexpected time, but they also made more errors. For both RT and IES, there

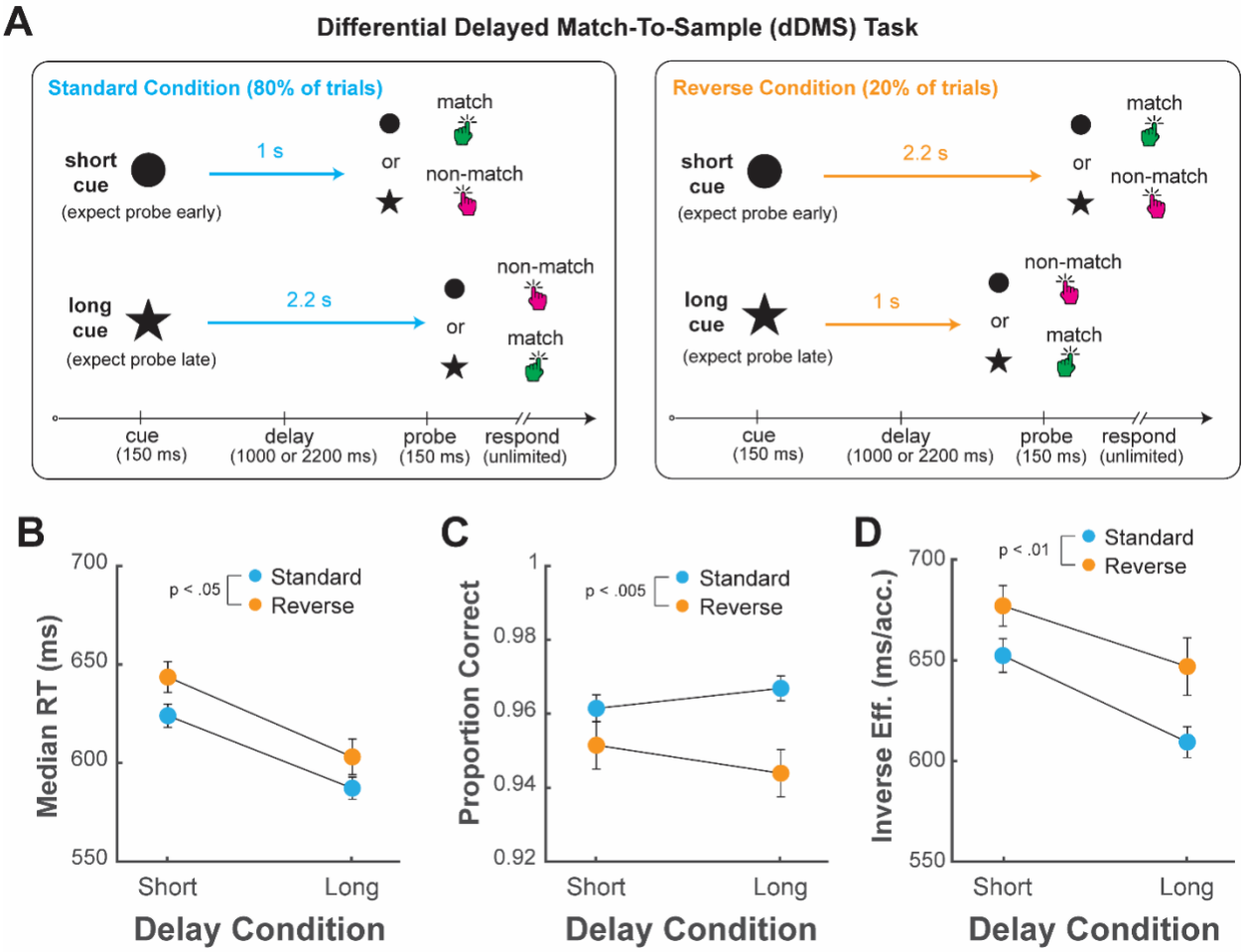

Figure 4.1 Timing biases performance in a working memory task with a cue-delay contingency. **A.** The differential-Delayed-Match-To-Sample (dDMS) task was based on a standard two-alternate forced choice implementation of delayed match-to-sample task, in which participants pressed a button to indicate whether two stimuli separated by a memory delay were the same (match) or different (non-match). However, cue identity predicted the duration of the delay period on 80% of the trials (Standard), while the cue-delay contingency was reversed on the remaining 20% of the trials (Reverse). The mapping between the cue stimulus and the likely memory delay was counterbalanced across participants. **B.** Median reaction time (RT) as a function of Delay and Reversal conditions. Median RT was significantly slower for the Reverse trials, and it was also significantly slower at the Short Delay. **C.** Accuracy as a function of Delay and Reversal conditions. Participants made significantly more errors on Reverse trials. **D.** Inverse efficiency (median RT / accuracy) as a function of Delay and Reversal conditions. Inverse efficiency is a measure which combines both speed and accuracy of performance and in which larger values indicate worse performance. Participants were significantly less efficient on Reverse trials, and they were also significantly less efficient at the Short Delay.

was also a main effect of Delay in which participants responded more quickly at the Long delay (RT: $F_{1, 26} = 30.7$, $p < .001$; IES: $F_{1, 26} = 14$, $p < .001$). Faster responses at the longer delay likely reflect the fact that since the probe could only appear at two potential delays, once the short delay had elapsed the participant could be more certain that it would appear at the long delay, which I refer to as the hazard rate effect [18].

To increase the validity of the results, I then performed a replication study of the dDMS task in an independent sample of participants (**Fig. S4.1**). In the replication study, there was a significant main effect of Reversal for both RT and IES (RT: $F_{1, 38} = 9.0$, $p < .005$; IES: $F_{1, 38} = 8.5$, $p < .01$) but not accuracy ($F_{1, 38} = 1.55$, $p = .22$). There was however a significant interaction between Reversal and Delay for accuracy ($F_{1, 38} = 4.1$, $p < .05$), and a significant simple effect of Reversal at the Long Delay ($t_{(38)} = 2.6$, $p < .05$), indicating that in the replication study participants made more errors when the common cue-delay contingency was reversed at the Long Delay. Just as in the original study, in the replication study there were significant effects of Delay for both RT and IES (RT: $F_{1, 38} = 31.3$, $p < .001$; IES: $F_{1, 38} = 36.1$, $p < .001$), reflecting the hazard rate effect. When Study (Original vs Replication) was added as a third factor in a statistical analysis that included all of the data, all Study-related factors were insignificant for all three measures (all $p > .05$), indicating that the two studies were statistically indistinguishable. Therefore, I also collapsed the two studies together and found that the pattern of significant factors was the same as in the original study, with significant main effects of Reversal for all three measures (RT: $F_{1, 65} = 16.6$, $p < .001$; accuracy: $F_{1, 65} = 9.0$, $p < .005$; IES: $F_{1, 65} = 17.5$, $p < .001$) and significant main effects of Delay for RT and IES (RT: $F_{1, 65} = 60.1$, $p < .001$; IES: $F_{1, 65} = 49.2$, $p < .001$).

4.2.2 The interval stimulus association (ISA) task

If timing and WM share coding mechanisms when memory delays are predictable, then one may predict that not only should timing information affect WM performance but that an inverse effect from WM onto timing should exist. In the dDMS task, which explicitly required WM, participants implicitly learned and were biased by task-irrelevant timing information. This motivates the intriguing question of whether in an explicit timing task, participants could implicitly learn and be biased by task-irrelevant WM information. To address this possibility, I developed an explicit timing task with a cue-delay contingency that I term the interval stimulus association task (ISA; **Fig. 4.2A**). The ISA task had the same stimulus structure as the dDMS task but differed in the task instructions and the correct response on Reverse trials. Specifically, participants were instructed to respond based on the combination of the duration of the delay and the identity of the probe. For one pair of opposing delay and probe stimulus combinations (e.g. short-circle and long-star), participants were required to press one button, while for the other pair of delay-probe combinations (e.g. short-star and long-circle), they were required to press a different button. Just as in the dDMS task, the cue predicted the delay duration in 80% of the trials, while for the other 20% of the trials the delay was the opposite duration. But unlike the dDMS task, correct responses were explicitly tied to the duration of the delay and so were inverted for Reverse trials. Importantly, participants were instructed that on any given trial the interval to be judged would be initiated by either a circle or a star and that the cue identity was irrelevant to the task beyond indicating the onset of the interval.

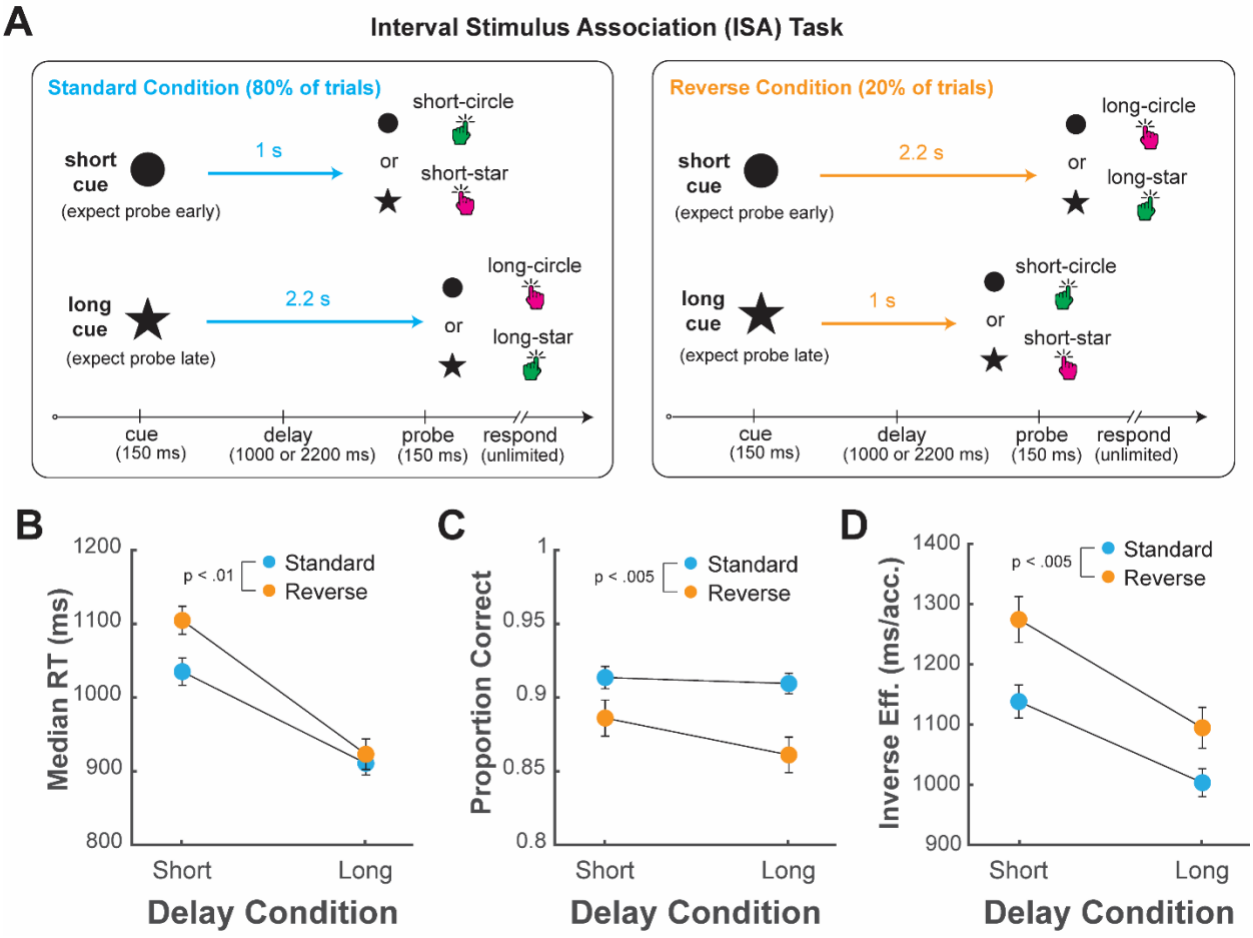

Figure 4.2 Cue identity biases performance in a timing task with a cue-delay contingency. **A.** In the Interval Stimulus Association (ISA) task, participants pressed a button based on the duration of the delay between the cue and probe stimuli as well as the identity of the probe stimulus. Each correct response mapped to two opposing delay-probe combinations (i.e. one response mapped to Short-Circle and Long-Star, while the other response mapped to Long-Circle and Short-Star). However, cue identity predicted the duration of the delay period on 80% of the trials (Standard), while the cue-delay contingency was reversed on the remaining 20% of the trials (Reverse). Unlike the dDMS task, because the delay was explicitly tied to the correct response, the correct response was also inverted on Reverse trials. The mapping between the cue stimulus and the likely cue-probe interval was counterbalanced across participants. **B.** Median reaction time (RT) as a function of Delay and Reversal conditions. Median RT was significantly slower for the Reverse trials, and it was also significantly slower at the Short Delay. **C.** Accuracy as a function of Delay and Reversal conditions. Participants made significantly more errors on Reverse trials. **D.** Inverse efficiency (median RT / accuracy) as a function of Delay and Reversal conditions. Inverse efficiency is a measure which combines both speed and accuracy of performance and in which larger values indicate worse performance. Participants were significantly less efficient on Reverse trials, and they were also significantly less efficient at the Short Delay.

For all three behavioral measures (RT, accuracy, and IES), there was a main effect of Reversal in which behavioral performance was worse for the Reverse trials (RT: $F_{1, 21} = 9.2$, $p < .01$; accuracy: $F_{1, 21} = 12.9$, $p < .005$; IES: $F_{1, 21} = 11$, $p < .005$), indicating that violation of the cue-delay association impaired timing-discrimination performance (**Fig. 4.2B-D**). In other words, participants made more errors and responded more slowly when the cue stimulus that initiated the delay period predicted a different time interval, which implies that participants maintained a working memory of the cue's identity until the arrival of the probe. For both RT and IES, there was also a main effect of Delay in which participants made their responses more quickly at the Long delay (RT: $F_{1, 21} = 25.7$, $p < .001$; IES: $F_{1, 21} = 12.5$, $p < .005$), reflecting the hazard rate effect.

To increase the validity of the results, I performed a replication study of the ISA task in an independent sample (**Fig. S4.2**). The pattern of significant factors in the replication study was the same as in the original study, with significant main effects of Reversal for all three behavioral measures (RT: $F_{1, 24} = 9.3$, $p < .01$; accuracy: $F_{1, 24} = 7.8$, $p < .05$; IES: $F_{1, 24} = 8.8$, $p < .01$) and significant main effects of Delay for both RT and IES (RT: $F_{1, 24} = 14.2$, $p < .001$; IES: $F_{1, 24} = 5.2$, $p < .05$). When Study (Original vs Replication) was added as a third factor in a statistical analysis that included all of the data, all Study-related factors were insignificant for all three measures (all $p > .05$), indicating that the two studies were statistically indistinguishable. I also collapsed the two studies together and found that the pattern of significant factors was the same, with significant main effects of Reversal for all three measures (RT: $F_{1, 46} = 18.5$, $p < .001$;

accuracy: $F_{1, 46} = 19.4$, $p < .001$; IES: $F_{1, 46} = 19.7$, $p < .001$) and significant main effects of Delay for RT and IES (RT: $F_{1, 46} = 37.9$, $p < .001$; IES: $F_{1, 46} = 15.9$, $p < .001$).

Together, the results of the dDMS and ISA tasks provide human behavioral evidence of a link between timing and WM, at least when the duration of the memory delay is predictable. In the dDMS task, participants explicitly encoded the cue stimulus identity into WM to perform the task but implicitly tracked the task-irrelevant delay. On the other hand, in the ISA task, participants explicitly tracked elapsed time after the cue to perform the task, but the results show that they also implicitly stored the task-irrelevant cue identity in WM. These results are consistent with the notion that WM and timing are multiplexed and stored in time-varying patterns. Ongoing computational studies in my lab support the notion that a strong candidate of this time-varying pattern is a cue-specific neural sequence. In such a coding scheme, each cue triggers a neural sequence that can be decoded to extract information not only about how much time has passed since the cue was presented but also its identity.

4.3 Discussion

Transiently storing past information and prospectively anticipating future events are widely considered to be among the most fundamental functions the brain performs [2-4, 19, 20]. Despite this, the fields of working memory (WM) and timing have developed largely independently of each other, as they have been generally considered separate mental functions with distinct neural implementations. However, on a computational level, timing and WM share similar functional constraints: both require transiently storing information – retrospective information in the case of WM and prospective information in the case of timing (e.g. when an event will occur in relation to

a previous stimulus). These properties can be viewed as intertwined, in much the same way that an hourglass both encodes the memory that it was recently turned over and also tells when a future event will take place.

4.3.1 Background

For years experimental psychologists have employed behavioral tasks with a rigidly repeating temporal structure, and the tradition of analyzing how temporal structure affects performance, primarily reaction time (RT), is now over sixty years old[21, 22]. One main conclusion has been that as the delay between a warning signal and a stimulus that requires a response becomes more predictable (i.e. has a smaller variance), participants respond more quickly[22]. This result alone suggests that the brain attempts to learn the temporal structure of external events to optimize performance. As the field of implicit timing took shape, researchers found that when cue *identity* predicts the duration of the delay, and the participants simply needed to press a button in response to target stimulus that appeared after a delay, RT is systematically faster when the response stimulus appears at the time predicted by the cue and systematically slower when the response stimulus appears at the time not predicted by the cue[23-25]. In this case, not only is the temporal structure of the task learned, but the cue's identity becomes in some sense bound to the corresponding temporal prediction, i.e. a cue-delay association.

Despite progress in understanding temporal prediction in simple target detection tasks, it has remained less clear how representations that underlie temporal prediction are incorporated into WM. Recently, a few studies have shown that when the duration of a WM delay is predictable, expected timing affects not only RT but also accuracy of the

memory judgement[13-15]. A leading conceptualization of these effects is that it reflects the allocation of *attention* in the temporal domain[19]. In the current study, I hypothesized that expected timing was constituted by a cue-specific temporal pattern of neural activity that simultaneously represented the cue identity and when the probe would likely appear. Based on previous and ongoing computational work in my lab[6], I specifically conjecture that the most likely form of the temporal pattern is a neural sequence, but other time-varying activity patterns such as ramping firing rates are also possible[12, 20]. However, the neural sequence hypothesis is compatible with the temporal attention hypothesis if one assumes that temporal attention is supported by the neural sequence (e.g. attention is increased at the end of the sequence).

4.3.2 Relationship between the hazard rate effect and the current results

A consistent finding across many prior studies is that when multiple delays are possible, participants respond more quickly following the longer delay regardless of cue-based temporal expectation[18]. This is because once the shorter delay has elapsed, the participant should be more certain that the target will appear at the longer delay, which is referred to as the hazard rate effect. Consistent with previous studies, I also observe the standard hazard rate effect in both the dDMS and ISA tasks (**Fig. 4.1B & 4.2B**).

For studies in which the cue predicts the likely timing of the target, the hazard rate effect has also been used to explain why for some studies the RT difference resulting from violation of cue-based expectations is asymmetric at the short and long delays, with large increases in RT when targets appear unexpectedly early and little to no difference when targets appear unexpectedly late[23, 25]. In the latter case cue-

based temporal expectations are at odds with the temporal prediction generated by the hazard rate, and it seems that the hazard rate can sometimes override cue-based temporal expectations[23, 25]. Although findings have been mixed as to whether cue-based expectation effects on RT are symmetric at early and late delays, most early studies of implicit timing that used simple target detection paradigms found smaller effects at the long delay[25], and some studies found no effect whatsoever at the long delay[23].

In the dDMS, I found that the RT deficit incurred by violating the cue-based expectation was roughly symmetrical at both the short and long delays (**Fig. 4.1B**). Among the three previous studies that used memory delay contingencies in a WM task that I am aware of, two also found symmetrical cue-based expectation effects on RT at short and long delays[13, 14], while the other did not find a negative effect of unexpected timing on RT at the long delay [15]. However, in the latter study's task structure, cue stimulus features were not associated with the likely delay; instead, temporal expectations came about because each block of trials predominantly used one memory delay, meaning that the temporal expectation was associated more with the context than the cue[15]. Thus I speculate that in the first two studies referenced above[13, 14] and in the dDMS task, expectation effects on RT are found at both short and long delays because the cue information maintained in WM provides a temporal prediction that conflicts with the hazard rate. As for the ISA task, the RT deficit caused by violating cue-based expectation was observed to be substantially smaller at the late delay in the initial sample of the ISA task (**Fig. 4.2B**), although statistically I only observed a main effect of delay and no interaction between cue and delay. However, in

the replication sample and in the combined original and replication samples, the RT effects appear to be nearly symmetrical (**Fig. S4.2A & S4.2D**). I speculate that unlike the hazard rate, cue-based temporal expectations affect both speed and accuracy such that the two may trade off[17], and as a result the IES measures generally show a more symmetrical effect of cue-based temporal expectation (**Fig. 4.1D & Fig. 4.2D**).

Another interesting feature of the current results is that although the hazard rate effect is beneficial for speed of response at the long delay, it does not appear to provide any benefit for the accuracy of the memory judgement in the dDMS task or the timing judgement in the ISA task (**Fig. 4.1C & 4.2C**). This result is consistent with the three previous studies that used WM delay contingencies[13-15]. The lack of a hazard rate effect on accuracy suggests that speed benefits at the long delay result from a distinct process, and the cause of errors cannot simply be attributed to generalized temporal surprise.

Just as for RT, cue-based expectation effects on accuracy are symmetrical at the short and long delays (**Fig. 4.1C & 4.2C**), which suggests that they result from the same process that slows RT when cue-based expectations are violated. The two previous studies in which cue features were associated with the likely memory delay found that the effect of violating cue-based expectations was roughly symmetrical at the short and long delays[13, 14], while the study that used context-based expectations found no difference in accuracy at the long delay[15]. Together, these results suggest that erroneous responses were made because at either delay participants used information both about cue identity (which was irrelevant in the ISA task) and probe timing (which

was irrelevant in the dDMS task), which I attribute to the employment of cue-specific time-varying representations.

4.3.3 Interpreting the results of the ISA task

The development of the ISA task was driven by our core hypothesis that timing and WM are inherently linked. Although previous studies had shown that task-irrelevant timing information affects WM performance [13, 15], a reciprocal effect in which WM information could affect timing performance had not been demonstrated. The results of the ISA task thus solidify evidence of the link between timing and WM by testing it from the other side, and to my knowledge, the ISA task is the first of its kind.

Interpreting the results of the ISA task is more difficult than the dDMS task, and because of the novelty of this task comparable studies are nonexistent to my knowledge. Studies of explicit timing typically follow psychophysical protocols and require participants to discriminate time intervals that are much more similar in magnitude (e.g. 10% difference). Although the intervals compared in the current task were highly distinct, participants also needed to discriminate the probe stimulus and map the interval-probe combination onto the correct response. As a result, the ISA task was more difficult than the dDMS task, which is reflected in the longer reaction times and lower accuracy on average (**Fig. 4.1 & 4.2**). Nevertheless, participants were still able to learn the task and achieve an average accuracy of about 90%. Interestingly, due to the two-alternative forced choice structure of the task, participants had to map each pair of opposing interval-probe combinations to a single button (e.g. short-circle and long-star both mapped to a single button). A consequence of this is that if participants made a mistake, it cannot be determined whether they misjudged the timing component

or the probe discrimination component (e.g. relative to short-circle, both short-star and long-circle are incorrect). However, given that categorizing the probe stimulus should have been trivial, errors likely resulted from misjudgment of the timing component or a more generalized cognitive conflict. Given the distinctiveness of the two possible delay durations, it is somewhat surprising that cue identity caused participants to misjudge the duration more frequently in trials with the reversed cue-delay relationship (**Fig. 4.2C**). In my view, this is because participants exploited the informativeness of the cue to aid in their temporal judgements, which underscores how prospective usage of cue information inherently served to form a WM of cue identity in the ISA task.

4.3.4 Future work and conclusions

Overall, the results of the dDMS and ISA tasks are consistent with my hypothesis that participants employ cue-specific time-varying representations that both memorize the cue identity and tell when the probe stimulus will likely appear. Future studies should directly test this idea by measuring neural activity in nonhuman animals during tasks with the same structure. I predict that population neural activity in areas such as hippocampus, striatum, or frontal cortex will exhibit cue-specific neural sequences[5-7] whose length matches the predicted probe timing. Another potential approach for understanding the mechanisms underlying timing and WM is to use a computational model and train the model neural network on the same tasks. If training biologically plausible models to perform the dDMS and ISA tasks causes the emergence of cue-specific neural sequences, it would provide strong evidence that cue-specific neural sequences serve as the computational link between timing and WM.

4.4 Methods

4.4.1 Participants

Across four experiments, a total of 130 human volunteers (62 female, 5 left-handed, mean age = 29, range 18-40) participated in the study. Data from seventeen participants were excluded from analysis due to low accuracy (less than 70%) or consistently slow reaction time (RT) such that too few trials met inclusion criteria (less than 50% of possible trials in any Reversal x Delay condition remaining after RT exclusion). All participants provided informed consent before participating and were paid for their participation. The study was approved by the Institutional Review Board of UCLA.

4.4.2 Online experimentation and recruitment

Experiments were conducted online, with hosting provided by Gorilla (https://gorilla.sc/) and recruitment provided by Prolific (https://www.prolific.co/). The precision and accuracy of timing on the Gorilla platform (i.e. of visual presentation and RT) has been studied in detail[26, 27]. Although RT measurements are inaccurate, being ~80 ms larger than the true RT, they are relatively precise with standard deviations of approximately 8-21 ms depending on the exact browser, operating system, and device[27]. Participants accessed the experiment using personal computers running Google Chrome or Mozilla Firefox. No other device types (i.e. phones or tablets) or browsers were allowed. Participants on the Prolific platform were only eligible for the study if they were between the ages of 18 and 40, residing in the United States, fluent in English, and had never participated in an online study from my laboratory on Prolific. Before beginning the task, participants read and signed an informed consent

form that asked them to: 1) complete the study in a quiet place without distractions, 2) maximize their browser window and not adjust it during the experiment, 3) have normal or corrected-to-normal vision (i.e. to wear glasses or contacts if prescribed), and 4) not participate if they had a history of seizures, epilepsy, or stroke. After providing consent, participants completed a short demographics form including their age, handedness, and gender. Participants were then given instructions on how to perform the task, which stressed the importance of both speed and accuracy. Participants were also informed that if they were faster and more accurate than the average of the other participants in a given sample of participants, they would receive a bonus payment.

4.4.3 The differentially delayed match-to-sample (dDMS) task

Participants performed a two-alternate forced choice task, which I term the differentially delayed match-to-sample (dDMS) task, in which I manipulated temporal expectation regarding expected probe time based on the encoded item (**Fig. 4.1A**). The background was always white, and all stimuli were black and presented in the center of the screen. First, a 150 ms duration fixation cross was presented, which indicated the start of a few trial. Following a 500-1000 ms interval, a 150 ms duration visual cue was presented, which could either be a black circle or black star, matched for area, with 50% probability. After a variable duration delay, a 150 ms duration probe stimulus was presented that was either the exact same stimulus or the opposite stimulus with 50% probability. Participants were instructed to press one of two buttons on their keyboards, F or J, to indicate whether they thought the cue and probe stimuli matched or did not match. The mapping between the response button and its meaning was counterbalanced across participants; for half of the participants, F indicated a match

and J a nonmatch, while this mapping was reversed in the other half. The response period was unlimited in duration, and the task did not proceed unless a response was given. All incorrect responses were followed by negative feedback (a “thumbs down” icon). After a response was given, there was a 1500-2000 ms inter-trial interval.

The critical manipulation in the current study involved the expected probe time. When appearing as a cue, one stimulus (e.g. the circle) was followed by a memory delay of 1000 ms on 80% of the trials and a delay of 2200 ms on the remaining 20% of the trials. In contrast, the other stimulus (e.g. the star) was followed by a 2200 ms delay on 80% of the trials and a 1000 ms delay on 20% of the trials. In this way, I expected participants to learn that one stimulus was likely to be a followed by a short delay, and the other was likely to be followed by a long delay. Trials in which the true delay was the most likely delay are referred to here as “Standard,” and trials in which the true delay was not the most likely delay are referred to as “Reverse.” The mapping between the cue stimulus and the likely memory delay was counterbalanced across participants.

Five blocks of 80 trials (64 valid, 16 invalid) were presented for a total of 400 trials. Blocks were identical in composition but had different trial orders. Trial order was pseudorandomized independently for each block for each participant with the following constraints: 1) The first eight trials of each block were always Standard. 2) A Reverse trial could not immediately follow another Reverse trial. Participants were also given eight valid training trials with each cue before the five experimental blocks. Participants were given the opportunity to take short breaks between each block. Each block took approximately eight minutes to complete, and participants finished the experiment in 45 minutes on average. At the end of the experiment, participants were asked whether they

noticed that one stimulus tended to be followed by a certain delay and provided a response of “Yes,” “No,” or “I don’t know.”

4.4.4 The interval-stimulus association (ISA) task

The interval-stimulus association (ISA) task was identical in stimulus structure to the dDMS task, but the correct responses and instructions were different (**Fig. 4.2A**). Rather than being instructed to compare the cue and probe stimuli to each other, participants were instructed to discriminate the time interval of the delay period and the probe stimulus. As described in the previous section, the delay interval could either be 1000 ms (Short) or 2200 ms (Long), and the probe stimulus could either be a Circle or a Star. Participants were instructed to press one key in response to one pair of opposing interval-probe combinations (e.g. press the F key in response to Short-Circle or Long-Star) and to press the other key in response to the other interval-probe combinations (e.g. press the J key in response to Long-Circle or Short-Star). The mapping between the response button and the pair of opposing interval-probe combinations was counterbalanced across participants (i.e. in the other half of participants, Short-Circle/Long-Star was associated with the J key, and Long-Circle/Short-Star was associated with the F key). Participants were instructed that for any given trial the cue stimulus could be either a Circle or Star and that the cue stimulus was thus irrelevant to the task beyond indicating the onset of the delay interval. But just as in the dDMS task, the cue stimulus identity (Circle or Star) predicted the delay interval on 80% of the trials (Standard trials), while for the remaining 20% of the trials, the cue stimulus was followed by the other delay interval (Reverse trials). However, in the ISA task, the correct response changed when the delay interval changed.

4.4.5 Analysis of Behavioral Data

Training trials were excluded from all analyses. Trials with RT outside of the range of 100-3000 ms were discarded. For the remaining trials, I considered three measures of performance: accuracy, RT, and the inverse efficiency score (IES). For accuracy, I calculated the proportion of trials of a given type with a correct response. For RT, I first discarded incorrect trials. Then, trials with outlying RT were discarded in the following way: For each participant and each Reversal x Delay condition separately, trials that were further than 4 standard deviations away from the mean were discarded. RT was calculated as the median of the remaining trials for that condition. Finally, the inverse efficiency score (IES), a combined measure of speed and accuracy in which larger values indicate worse performance, was calculated as follows:

$$IES = \frac{median\ RT}{accuracy}$$

4.4.6 Statistical analysis

There were no between-subjects factors. Trials were classified based on Reversal (Standard vs. Reversal) and Memory Delay (Probe Early vs. Probe Late). Thus, I used a 2 x 2 repeated-measures ANOVA with factors of Reversal and Memory Delay and used follow-up paired-samples *t* tests to test for Reversal effects at each Memory Delay separately when there was an interaction.

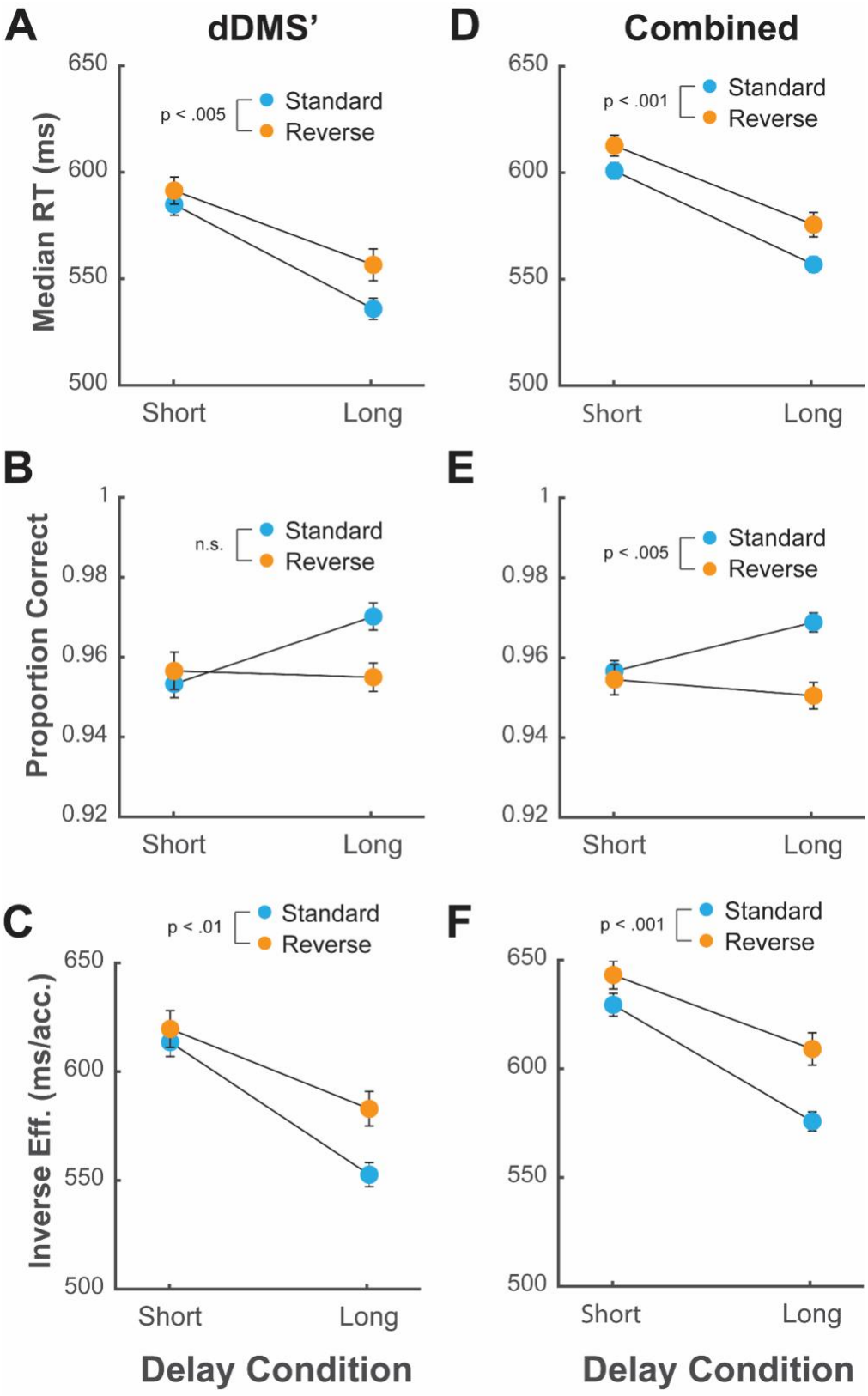
A
dDMS'
D
Combined
Median RT (ms)
650
600
550
500
p < .005
Standard
Reverse
Short
Long
p < .001
B
E
Proportion Correct
1
0.98
0.96
0.94
0.92
n.s.
C
F
Inverse Eff. (ms/acc.)
p < .01
Delay Condition

Figure S4.1.

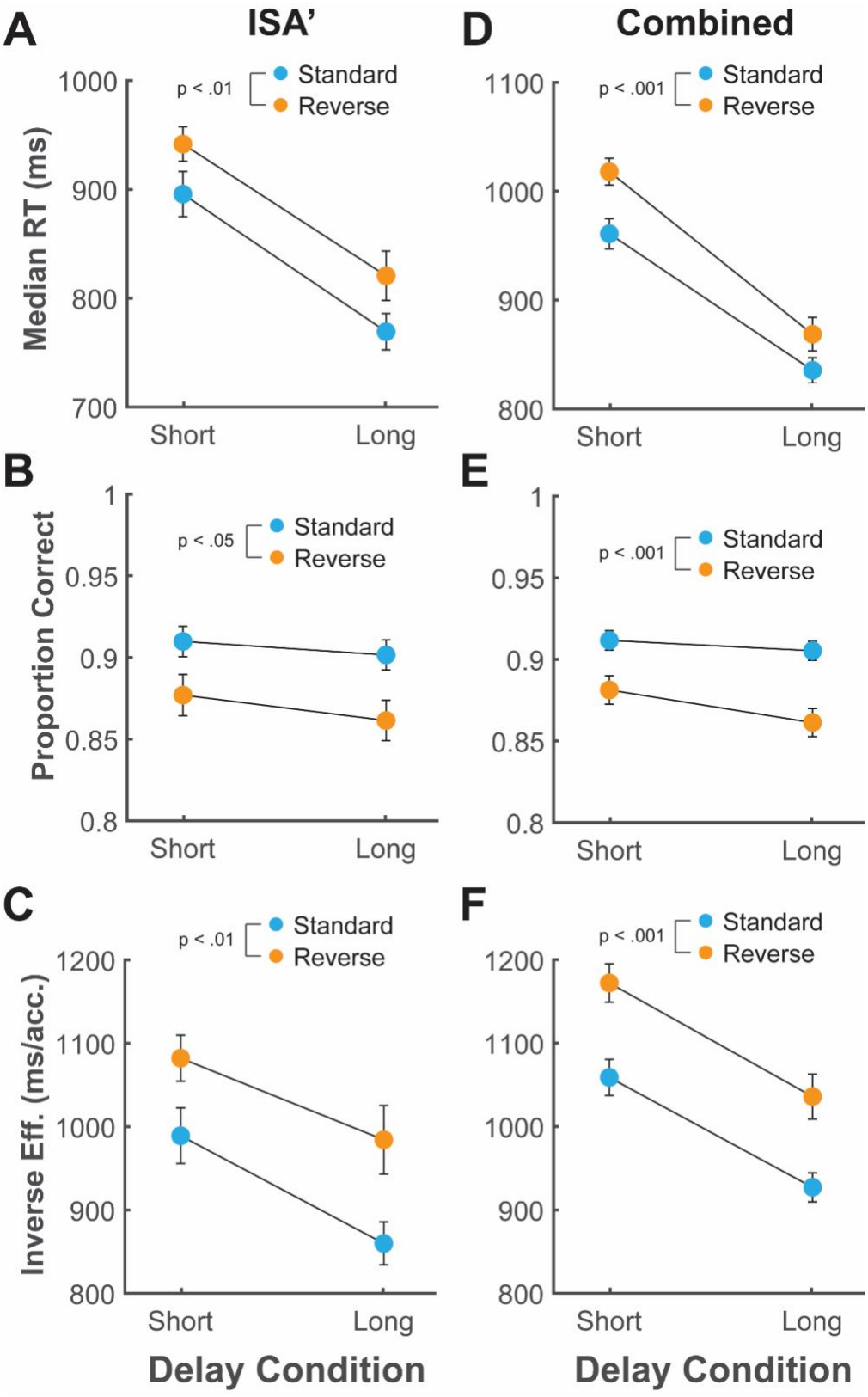
A
ISA'
Median RT (ms)
1000
900
800
700
p < .01
Standard
Reverse
Short
Long
D
Combined
1100
1000
900
800
p < .001
Standard
Reverse
Short
Long
B
Proportion Correct
1
0.95
0.9
0.85
0.8
p < .05
Standard
Reverse
Short
Long
E
1
0.95
0.9
0.85
0.8
p < .001
Standard
Reverse
Short
Long
C
Inverse Eff. (ms/acc.)
1200
1100
1000
900
800
p < .01
Standard
Reverse
Short
Long
Delay Condition
F
1200
1100
1000
900
800
p < .001
Standard
Reverse
Short
Long
Delay Condition

Figure S4.2.

Figure S4.1 Timing biases behavioral performance in an independent replication sample of the dDMS.
A. In an independent replication sample of the dDMS task, median reaction time (RT) on correct trials was slower when the more common relationship between the cue and the working memory delay (80%) was reversed (20%). median RT on correct trials was slower when the more common relationship between the cue and the working memory delay (80%) was reversed (20%). In addition to the significant main effect of Reversal, there was a significant main effect of Delay, indicating a hazard rate effect. **B.** In the dDMS replication study, participants were not less accurate on average when the common cue-delay relationship was reversed. However, there was a significant interaction between Reversal and Delay qualified by a simple effect of Validity at the Long Delay. **C.** In the dDMS replication study, participants had worse (larger) inverse efficiency scores when the common cue-delay relationship was reversed**.** In addition to the significant main effect of Reversal, there was a significant main effect of Delay, indicating a hazard rate effect. **D.** In a sample that combined the original and replication study samples for the dDMS task, median RT on correct trials was slower when the common cue-delay relationship was reversed. In addition to the significant main effect of Reversal, there was a significant main effect of Delay, indicating a hazard rate effect. **E.** In the combined sample, participants were less accurate on average when the common cue-delay relationship was reversed. For accuracy, there was only a significant main effect of Reversal. **F.** In the combined sample, participants had worse (larger) inverse efficiency scores when the common cue-delay relationship was reversed. In addition to the significant main effect of Reversal, there was a significant main effect of Delay, indicating a hazard rate effect.

Figure S4.2 Cue identity biases behavioral performance in an independent replication sample of the ISA.
A. In an independent replication sample of the ISA task, median reaction time (RT) on correct trials was slower when the more common relationship between the cue and the working memory delay (80%) was reversed (20%). median RT on correct trials was slower when the more common relationship between the cue and the working memory delay (80%) was reversed (20%). In addition to the significant main effect of Reversal, there was a significant main effect of Delay, indicating a hazard rate effect. **B.** In the ISA replication study, participants were less accurate on average when the common cue-delay relationship was reversed. For accuracy, there was only a significant main effect of Reversal. **C.** In the ISA replication study, participants had worse (larger) inverse efficiency scores when the common cue-delay relationship was reversed**.** In addition to the significant main effect of Reversal, there was a significant main effect of Delay, indicating a hazard rate effect. **D.** In a sample that combined the original and replication study samples for the ISA task, median RT on correct trials was slower when the common cue-delay relationship was reversed. In addition to the significant main effect of Reversal, there was a significant main effect of Delay, indicating a hazard rate effect. **E.** In the combined sample, participants were less accurate on average when the common cue-delay relationship was reversed. For accuracy, there was only a significant main effect of Reversal. **F.** In the combined sample, participants had worse (larger) inverse efficiency scores when the common cue-delay relationship was reversed. In addition to the significant main effect of Reversal, there was a significant main effect of Delay, indicating a hazard rate effect.

4.5 Chapter 4 References

1. Buhusi, C.V. and W.H. Meck, *What makes us tick? Functional and neural mechanisms of interval timing.* Nature Reviews Neuroscience, 2005. **6**(10): p. 755-765.

2. Merchant, H., D.L. Harrington, and W.H. Meck, *Neural Basis of the Perception and Estimation of Time.* Annual Review Neuroscience, 2013. **36**(May): p. 313-36.

3. Baddeley, A.D. and G. Hitch, *Working Memory*, in *Psychology of Learning and Motivation*, G.H. Bower, Editor. 1974, Academic Press. p. 47-89.

4. D'Esposito, M. and B.R. Postle, *The cognitive neuroscience of working memory.* Annu Rev Psychol, 2015. **66**: p. 115-42.

5. Bakhurin, K.I., et al., *Differential Encoding of Time by Prefrontal and Striatal Network Dynamics.* The Journal of Neuroscience, 2017. **37**(4): p. 854-870.

6. Zhou, S., S.C. Masmanidis, and D.V. Buonomano, *Neural Sequences as an Optimal Dynamical Regime for the Readout of Time.* Neuron, 2020. **108**(4): p. 651-658.e5.

7. Taxidis, J., et al., *Differential Emergence and Stability of Sensory and Temporal Representations in Context-Specific Hippocampal Sequences.* Neuron, 2020. **108**(5): p. 984-998.e9.

8. Petter, E.A., S.J. Gershman, and W.H. Meck, *Integrating Models of Interval Timing and Reinforcement Learning.* Trends in Cognitive Sciences, 2018. **22**(10): p. 911-922.

9. Brody, C.D., et al., *Timing and Neural Encoding of Somatosensory Parametric Working Memory in Macaque Prefrontal Cortex.* Cerebral Cortex, 2003. **13**(11): p. 1196-1207.

10. Stokes, M.G., *'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework.* Trends in Cognitive Sciences, 2015. **19**(7): p. 394-405.

11. Constantinidis, C., et al., *Persistent Spiking Activity Underlies Working Memory.* J Neurosci, 2018. **38**(32): p. 7020-7028.

12. Lundqvist, M., P. Herman, and E.K. Miller, *Working Memory: Delay Activity, Yes! Persistent Activity? Maybe Not.* J Neurosci, 2018. **38**(32): p. 7013-7019.

13. van Ede, F., M. Niklaus, and A.C. Nobre, *Temporal expectations guide dynamic prioritization in visual working memory through attenuated α oscillations.* Journal of Neuroscience, 2017. **37**(2): p. 437-445.

14. Zokaei, N., et al., *Modulation of the pupillary response by the content of visual working memory.* Proceedings of the National Academy of Sciences, 2019. **116**(45): p. 22802-22810.

15. Jin, W., A.C. Nobre, and F. van Ede, *Temporal expectations prepare visual working memory for behavior.* Journal of Cognitive Neuroscience, 2020. **32**(12): p. 2320-2332.

16. Townsend, J. and F. Ashby, *Methods of modeling capacity in simple processing systems*. 1978. p. 199-239.

17. Vandierendonck, A., *On the Utility of Integrated Speed-Accuracy Measures when Speed-Accuracy Trade-off is Present.* Journal of cognition, 2021. **4**(1): p. 22-22.

18. Nobre, A.C., A. Correa, and J.T. Coull, *The hazards of time.* Current Opinion in Neurobiology, 2007. **17**(4): p. 465-470.

19. Nobre, A.C. and F. Van Ede, *Anticipated moments: Temporal structure in attention.* Nature Reviews Neuroscience, 2018. **19**(1): p. 34-48.

20. Paton, J.J. and D.V. Buonomano, *The Neural Basis of Timing: Distributed Mechanisms for Diverse Functions.* Neuron, 2018. **98**(4): p. 687-705.

21. Foley, P.J., *The foreperiod and simple reaction time.* Canadian Journal of Psychology/Revue canadienne de psychologie, 1959. **13**(1): p. 20-22.

22. Niemi, P. and R. Näätänen, *Foreperiod and simple reaction time.* Psychological Bulletin, 1981. **89**: p. 133-162.

23. Miniussi, C., et al., *Orienting attention in time. Modulation of brain potentials.* Brain, 1999. **122**(8): p. 1507-1518.

24. Coull, J.T., et al., *Orienting attention in time: behavioural and neuroanatomical distinction between exogenous and endogenous shifts.* 2000. **38**: p. 1-12.

25. Nobre, A.C., *Orienting attention to instants in time.* Neuropsychologia, 2001. **39**(12): p. 1317-1328.

26. Bridges, D., et al., *The timing mega-study: comparing a range of experiment generators, both lab-based and online.* PeerJ, 2020. **8**: p. e9414-e9414.

27. Anwyl-Irvine, A., et al., *Realistic precision and accuracy of online experiment platforms, web browsers, and devices.* Behavior Research Methods, 2020.

Chapter 5: Conclusion

When things happen at different times, they become in a sense disconnected from each other. As a result, it's not trivial to extract meaningful information from of a pattern of events that plays out over time, and many organisms' abilities in this regard are deeply limited. Humans excel at parsing the temporal structure of experience, which allows us to recognize speech, detect regularity in sequences of events, and predict when things will happen.

Although humans can tell time across ten orders of magnitude[1], most sophisticated sensory temporal processing abilities take place on the scale of tens-to-hundreds of milliseconds. Naturally-occurring behaviors in nonhuman organisms demonstrate that temporal features on this scale such as interval, rate, and duration are critical for communication, courtship, and even navigation[2-4]. Accordingly, many organisms possess sensory neurons whose responses are selective to the behaviorally-relevant timing features[5-7]. More generally it has been shown that in a number of species, the sensory neuronal response to a stimulus depends on the recent history of inputs, which I refer to as temporal context[8, 9]. In the current work, I propose that neurons become endowed with sensitivity to timing and temporal context on the subsecond scale due to time-varying neuronal and synaptic properties, and perhaps the most useful tool for creating temporal selectivity on this scale is short-term synaptic plasticity (STP)[10].

In order to investigate the mechanisms underlying sensitivity to temporal context, I created a spiking model to explain the diversity of single pyramidal (Pyr) neuron responses to repeated tones, which included steady and facilitating responses in

addition to the well-studied phenomenon of short-term sensory adaptation[11]. The model incorporated two different types of inhibitory units based on Parvalbumin (PV) and Somatostatin (SST) inhibitory interneurons, and I modeled the STP of the synapses between Pyr, PV, and SST units based on empirical observations[12]. By varying the strength of the PV-to-Pyr and SST-to-Pyr synapses, my model could transition between adapting, steady, and facilitating temporal profiles of responses to repeated inputs. Furthermore, my model made two key predictions that were verified in the experimental data: 1) Neurons with steady firing rates across repeated stimuli should have shorter firing latencies than adapting or facilitating neurons, and 2) optogenetically inactivating PV interneurons during the first tone in a train of repeated tones should cause a decreased response to the second tone occurring 400 ms later. These results suggest that the way sensory responses are modulated by the temporal structure of a stimulus sequence depends on intra-cortical STP, and the modulation often results from STP-driven shifts in the firing latency of the neurons.

In contrast to sensory timing, many sophisticated forms of *internal* timing (e.g. temporal expectation or motor timing) take place on the scale of seconds. Short-term memory is another potentially mechanistically-related computation that also relies on internal neural network dynamics on this time scale. Short-term memory has been proposed to rely on stable persistent activity states generated by local recurrent neural networks[13]. In my work, I contributed to a growing body of theoretical work regarding the long-term plasticity rules that can configure synaptic weights of neural networks to support stable persistent activity states. My results show that inhibition-stabilized networks (ISNs)[14], in which strong recurrent excitation is balanced by inhibition,

provide a highly reliable synaptic regime to enable stable persistent activity states. However, ISNs necessitate learning rules that anticipate and deal with their unintuitive properties. In a large, sparsely-connected spiking model, I reinforce the recent finding that one solution is a family of “cross-homeostatic” learning rules, in which excitatory and inhibitory neurons alter their incoming weights according to the average input they receive from the opposite population[15]. Interestingly, I provide a novel finding that cross-homeostatic plasticity rules can operate using purely local signals, but only if classical homeostatic plasticity rules are also in place.

Because natural contexts often have predictable temporal structure, requirements to maintain information in working memory and to anticipate the timing of future events are often intermixed and overlapping. A growing body of experimental data suggests that there are similar patterns of neural activity during the maintenance of working memory and the anticipation of a time interval[16, 17], which suggests that the two functions may be computationally linked. To test this idea at the level of human behavior, I designed two tasks with the same predictable temporal structure but complementary behavioral requirements. The first task only required participants to maintain a cue stimulus in working memory, but the results showed that performance was worse when the probe stimulus occurred at an unpredicted time. The second task only required participants to judge the duration of the interval between the cue and probe stimuli, but the results showed that performance was worse when the cue identity predicted a different time interval, implying that participants stored cue information in working memory. This pattern of results is consistent with the notion that if the brain can predict *when* working memory will be used, it utilizes a shared coding format that fulfills

the functional requirements of both prospective timing and working memory maintenance. Based on prior experimental findings and ongoing computational work, I speculate that the most likely coding format is a cue-specific neural sequence, in which single neurons fire sparsely at specific times during the delay period, and the population activity can be used to decode both the content of working memory and also when a future event will take place.

As a whole, my work demonstrates that parsing the temporal structure of experience is a multifaceted challenge with distinct functional requirements depending on the timescale and behavioral pressures. In order to sense timing features on the scale of tens-to-hundreds of milliseconds, which enables for example the discrimination of phonemes and the perception of prosody in speech, the brain is likely to rely heavily on STP to create a rich representational substrate in which some neurons respond selectively to certain time intervals or durations and there is a mixture of adapting and facilitating responses in the population response. Yet in order to prospectively anticipate time intervals on the scale of seconds, the brain must instead rely on network mechanisms that can bring about persistent activity states or time-varying patterns such as neural sequences, whose durations are able to outlast the time constants of STP. In turn, the ability to generate persistent activity states and neural sequences likely depends on long-term plasticity mechanisms. Nervous systems thus stand as a cumulative history of the myriad evolved solutions for overcoming the challenge of time at various scales. In humans, these mechanisms form a stunning hierarchical symphony, allowing us to not only use temporal patterns as a medium for meaning but to orient ourselves within the temporal structure of events.

5.1 Chapter 5 References

1. Buonomano, D.V., *The biology of time across different scales*. 2007. p. 594-597.

2. Pollack, G., *Who, what, where? Recognition and localization of acoustic signals by insects.* Current Opinion in Neurobiology, 2000. **10**(6): p. 763-767.

3. Gerhardt, H.C. and F. Huber, *Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions*. 2002.

4. Covey, E. and J.H. Casseday, *Timing in the auditory system of the bat.* Annual Review of Physiology, 1999. **61**: p. 457-476.

5. Kostarakos, K. and B. Hedwig, *Calling Song Recognition in Female Crickets: Temporal Tuning of Identified Brain Neurons Matches Behavior.* Journal of Neuroscience, 2012. **32**(28): p. 9601-9612.

6. Rose, G.J. and R.R. Capranica, *Temporal selectivity in the central auditory system of the leopard frog.* Science, 1983. **219**(4588): p. 1087-1089.

7. Carlson, B.A., *Temporal-Pattern Recognition by Single Neurons in a Sensory Pathway Devoted to Social Communication Behavior.* Journal of Neuroscience, 2009. **29**(30): p. 9417-9428.

8. Asari, H. and A.M. Zador, *Long-lasting context dependence constrains neural encoding models in rodent auditory cortex.* Journal of Neurophysiology, 2009. **102**(5): p. 2638-2656.

9. Angeloni, C. and M.N. Geffen, *Contextual modulation of sound processing in the auditory cortex.* Current Opinion in Neurobiology, 2018. **49**: p. 8-15.

10. Motanis, H., M.J. Seay, and D.V. Buonomano, *Short-Term Synaptic Plasticity as a Mechanism for Sensory Timing.* Trends in Neurosciences, 2018. **41**(10): p. 701-711.

11. Seay, M.J., et al., *Differential Short-Term Plasticity of PV and SST Neurons Accounts for Adaptation and Facilitation of Cortical Neurons to Auditory Tones.* The Journal of neuroscience : the official journal of the Society for Neuroscience, 2020. **40**(48): p. 9224-9235.

12. Campagnola, L., et al., *Connectivity and Synaptic Physiology in the Mouse and Human Neocortex.* bioRxiv, 2021.

13. Wang, X.J., *Synaptic reverberation underlying mnemonic persistent activity.* Trends Neurosci, 2001. **24**(8): p. 455-63.

14. Sanzeni, A., et al., *Inhibition stabilization is a widespread property of cortical networks.* bioRxiv, 2019: p. 1-39.

15. Soldado-Magraner, S., et al., *Orchestrated Excitatory and Inhibitory Learning Rules Lead to the Unsupervised Emergence of Up-states and Balanced Network Dynamics.* bioRxiv, 2021: p. 2020.12.30.424888-2020.12.30.424888.

16. Taxidis, J., et al., *Differential Emergence and Stability of Sensory and Temporal Representations in Context-Specific Hippocampal Sequences.* Neuron, 2020. **108**(5): p. 984-998.e9.

17. Zhou, S., S.C. Masmanidis, and D.V. Buonomano, *Neural Sequences as an Optimal Dynamical Regime for the Readout of Time.* Neuron, 2020. **108**(4): p. 651-658.e5.