UNIVERSITY OF CALIFORNIA SAN DIEGO

Elucidating the role of retrosplenial cortex in history-based decision-making

A dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Neurosciences

by

Bethanny Patricia Danskin

Committee in charge:

Professor Takaki Komiyama, Chair
Professor Edward Callaway
Professor Timothy Gentner
Professor Christina Gremel
Professor John Serences

2022

The Dissertation of Bethanny Patricia Danskin is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

Dedication

To my mother for her relentless optimism, and my father for his unbounded curiosity.

EPIGRAPH

The only thing that makes life possible is permanent, intolerable uncertainty; not knowing what comes next.

Ursula Le Guin

TABLE OF CONTENTS

LIST OF FIGURES

List of Abbreviations

AIC	Akaike Information Criterion
ALM	Anterolateral motor cortex
CV	Cross-validated
loglik	Natural log of the likelihood
M1	Primary motor cortex
NMDA	N-Methyl-D-aspartate
pM2	Posterior secondary motor cortex
PPC	Posterior parietal cortex
Q	Value
RL	Reinforcement learning
RPE	Reward prediction error
RSC	Retrosplenial cortex
RW	Rescorla-Wagner
S1	Primary sensory cortex
V1	Primary visual cortex

Acknowledgements

I would like to offer deep gratitude to my mentor Takaki Komiyama for his support, encouragement, and scientific guidance throughout my doctoral education. He was consistently present and open to conversation on science and academia and life in general. He is a brilliant scientist with a talent for helping students develop a full complement of skills for the lab and beyond. In addition to offering his support, Takaki has built a lab environment full of smart, generous, hard-working people who all want the best for the lab and for each other.

I would specifically like to thank Ryoma Hattori, whose research was foundational to this dissertation, and whose work ethic is nothing short of inspirational. Ryoma is a dedicated scientist with a careful handling of both experiments and data analysis, and I have learned so much from working with him over the years. I would also like to thank Zeljana Babic, Nicole Mlynaryk, and Eva Zhang as other members of the Dynamic Foraging team whose input was critical to this project, and whose emotional support was invaluable. I would also like to thank other members of the Komiyama lab past and present, including especially Wankun Li, EunJung Hwang, Nathan Hedrick, Assaf Ramot, and Keelin O'Neil, whose expertise and advice and humor made lab life better even when experiments went awry.

I offer heartfelt gratitude and appreciation to Erin Gilbert and Linh Vandermar for their reassurance, encouragement, and endless patience for the students of the Neurosciences Graduate Program. You made navigating graduate school feel easy, and I can only imagine the work that went on behind the scenes.

I want to thank my fellow Neurosciences Graduate Program classmates and friends, who became my family away from home and are now my family across the world. I would especially like to thank my covid-quarantine team, the Surgebinders, who kept me sane and

healthy and connected during difficult times. And I state my profound appreciation for the Art of Espresso, the local campus coffee shop that made the best coffee and provided the most comforting location for all kinds of talks about science, life, joys and defeats.

Finally, I would like to thank my parents for their support and good humor for all these years. I would not be a fraction of who I am without your encouragement. I am so grateful for my brother and sisters and all the nieces and nephews who became little humans while I wasn't looking, and for my aunt and California cousins who endeavored to kidnap me away from lab when I needed it. Thank you all.

Chapter 2, in part, contains the reproduction of material as it appears in Hattori R, Danskin B, Babic Z, Mlynaryk N, and Komiyama T (2019). Area-Specificity and Plasticity of History-Dependent Value Coding During Learning. *Cell* 177, 1-15. The dissertation author was an author of this paper, and the primary researcher for included portion.

Chapter 3, in full, is material currently being prepared for submission for publication. Danskin B, Hattori R, Yu EZ, Aoi M, Komiyama T. Diverse behavioral timescales encoded in retrosplenial cortex explain hyperbolic behavior. The dissertation author was the primary author of this material.

Vita

2010 Associate of Arts and Sciences, Bellevue College

2013 Bachelor of Science in Neurobiology, University of Washington

2022 Doctor of Philosophy in Neurosciences, University of California San Diego

PUBLICATIONS

Hattori R, **Danskin B**, Komiyama T (2019). Area-specificity and plasticity of history-dependent value coding during learning. *Cell* 177, 1–15. 10.1016/j.cell.2019.04.027

Hedrick T, **Danskin B**, Larson RS, Ollerenshaw D, Groblewski PA, Valley M, Olsen S, & Waters J (2016). Characterization of channelrhodopsin and archaerhodopsin in cholinergic neurons of Cre-lox transgenic mice. *PLoS One*, 11(5)e0156596.

Danskin B, Denman D, Valley M, Ollerenshaw D, Williams D, Groblewski PA, Reid RC, Olsen S & Waters J (2015). Optogenetics in mice performing a visual discrimination task: measurement and elimination of retinal activation and the resulting behavioral artifact. *PLoS One*, 10(12):e0144760.

AWARDS

2018 Achievement Rewards for College Scientists Foundation Scholar, San Diego Chapter

2018 National Institutes of Health, F31 Ruth L. Kirschstein National Research Service Award Predoctoral Fellowship

2013 University of Washington's President's Medal

2013 Mary Gates Research Scholarship

2012 Computational Neuroscience Undergraduate Training Grant

ABSTRACT OF THE DISSERTATION

Elucidating the role of retrosplenial cortex in history-based decision-making

by

Bethanny Patricia Danskin

Doctor of Philosophy in Neurosciences

University of California San Diego, 2022

Professor Takaki Komiyama, Chair

Using past experience to inform future choice is fundamental to decision-making and behavior. Integrating experience by comparing choices and outcomes across time is a critical task, and requires a neural mechanism by which information may be assessed and accumulated. This is a widespread phenomenon in the brain, involving many cortical and subcortical structures. In this dissertation I show that one cortical area, the retrosplenial cortex (RSC), is particularly enriched in neurons that encode behaviorally-relevant history information, and is necessary for decision-making that relies on reward-history. RSC neurons exhibit a diversity of time-constants over which this history information is integrated, and I

have found that the timescales encoded in RSC match the temporal characteristics of the behavior better than other cortical areas. I developed a novel behavioral model in which decision is reached as the weighted sum of multiple exponential integrators using the observed diversity of time-constants. Acutely inactivating RSC results in the attenuation of this combinatorial behavioral strategy, and a decreased reliance on reward-history. From these results, I propose a conceptual model where reward-history information is encoded in neurons with a simple update rule, but the time-constants are heterogenous and vary across the population. The combination of diverse temporal information produces a behavioral strategy which is sensitive to both recent experience and long-term trends, a feature observed as the hyperbolic discounting of past experience.

In Chapter 1 I introduce the concepts of reinforcement learning theory relevant for this dissertation, and survey how reward-based value information is encoded in the brain. Chapter 2 identifies RSC as particularly important for the integration of past reward experience into actionable value information. Chapter 3 further examines and tests the role of RSC in integrating information across a diversity of timescales, and proposes a model of independent temporal integration in the brain that underlies the hyperbolic discounting of past experience. Chapter 4 discusses the properties of RSC that support the integration and maintenance of diverse information, and contextualizes the results in the broader decision-making context.

CHAPTER 1: INTRODUCTION

When we are presented with two or more options—as we are in many ways, many times throughout our day—we need to evaluate those options and decide which aligns closest to our interests. In this process we use our past experience to guide future choice. To do this we, and all animals, need a mechanism by with to integrate our recent experience in order to evaluate what is the better option. Choosing between options on the basis of the rewarded outcomes associated with those options is called value-based decision-making.

Sometimes these choices have clear and obvious outcomes associated with them, which can be intuited with little or no learning required. More often, however, we are presented with choices that have ambiguity in their probability of a positive outcome, and therefore value is something that must be learned through trial and error. As a concrete example, consider choosing between two coffee shops at which to indulge in a morning latte habit. There are many factors that might influence that choice: the quality of coffee, the speed of service, the distance traveled to and from the coffee shop, and the busyness at certain times of day. There are also factors that might change unpredictably across time, such as a change in supplier, limited seasonal selection, or the departure of a favorite barista. Each of these changes to the overall value associated with either coffee shop occurs with some different underlying timescale and frequency, and so when comparing options it is beneficial to simultaneously recall and integrate experiences across a range of time. Recent experience is especially salient, but long-term expectation setting is also beneficial.

How and why an animal chooses the action it does is fundamental to understanding behavior across a wide variety of disciplines, including neuroscience, psychology, economics, and medicine. Correspondingly, the tools to understand, describe, and predict behavior have

been drawn from a variety of mathematical models and theories. This dissertation will first introduce the concepts of reinforcement learning theory as foundational to the study of how and where decision-making is encoded in the brain. Then the author will detail the research identifying a candidate area, the retrosplenial cortex, as particularly relevant and necessary to decision-making based on the integration of past experience.

1.1 Reinforcement learning as a framework to understand decision-making

Reinforcement learning (RL) theory (Sutton and Barto 1998) provides a theoretical framework for learning and decision-making, and has had wide-ranging impact in both designing experiments and interpreting results. The advantage of this computational grounding is to attach numbers to alternative actions, such that internally derived representations of value and integrated experience may be treated quantitatively, and the choices can be understood as selecting an action that maximizes this quantity.

While there are a variety of implementations of RL algorithms for different applications, all share a common feature: an internally maintained value function that is updated according to experience of choice and outcome pairings. A learner cannot know the future, or even necessarily all the present dynamics underlying the probability of a rewarding stimulus. Thus, value functions reflect only the learner's best estimate of future rewards, which is necessarily contingent on its past experience. If the experienced outcome perfectly matches the learner's expected value, then there is no prediction error and no learning need occur. But in an environment where there is some stochasticity in the relationship between action and reward, the prediction may not match the observation. This creates an error, the difference between expected and observed, that is called reward prediction error, RPE (Sutton and Barto 1998). Given a quantification of expected value and a quantification of the

observed reward, the signed difference between these variables may be used to revise the expected value. Therefore, RPE is the primary driver of learning and is used to update the expected value for the chosen action.

In addition to describing many behavioral phenomena in psychology, cognitive science, and economics, RL models also provide access to the hidden variables of decision making (i.e. estimated value and RPE) that are not directly observable, but which nevertheless have correlates in the brain. Neuroscience research from humans (Daw et al. 2006; Behrens et al. 2007; Meder et al. 2017), non-human primates (Platt and Glimcher 1999; Sugrue 2004; Barraclough, Conroy, and Lee 2004; Samejima et al. 2005; Seo and Lee 2007; Lau and Glimcher 2008; Seo and Lee 2009; So and Stuphorn 2010; Cai, Kim, and Lee 2011; Massi, Donahue, and Lee 2018), and other animal models (Ito and Doya 2009; Kim et al. 2009; Sul et al. 2010; 2011; Hattori et al. 2019; Bari et al. 2019; Steinmetz et al. 2019) suggest that these computations are widespread, and carried out in multiple interconnected brain areas. However, the mechanisms by which the brain integrates past experience to shape future decision remains a developing and topical area of research.

1.2 The rate of learning sets the temporal window of history integration

The brain faces a difficult task in learning about the consequences of an action, namely how sensitive it should be to any individual outcome. If the chosen option has an outcome that is in some way noisy—varying in probability or magnitude—then being too sensitive to a single unrewarding experience may cause the animal to reject that option even through on average the option is rewarding. The rate of value update needs to be appropriate to the environment and the task demands in order to maximize rewards.

In the class of RL models called Q-learning (Sutton and Barto 1998), the value (Q) for a chosen action is updated according to the reward prediction error (RPE), with some gain controlled by a model parameter called the learning rate. A large learning rate indicates the learner adapts expectation quickly in response to new information, but necessarily loses sensitivity to more distant experience. This may be advantageous in an environment that is undergoing change, or in which the animal has a large amount of uncertainty about the potential outcomes of their choices (Behrens et al. 2007). Such unexpected uncertainty (Yu and Dayan 2005) is associated with volatile environments and large learning rates, that the animal may adapt more quickly.

However, large learning rates can be maladaptive in a stable environment if it is characterized by probabilistic or stochastic outcomes. Being overly sensitive to random fluctuations in recent outcomes might cost the animal the accumulated insight to the longer trends in the reward contingencies. The size of the learning rate is thus a manifestation of a bias-variance tradeoff.

1.3 Variability in temporal learning rates

Traditional formulations of Q-learning models include the learning rate as a single, fixed parameter. However, from a conceptual perspective, the advantage of having access to information across multiple temporal horizons is clear: real world situations change dynamically and unpredictably. Integrating across short timescales allows an animal to respond quickly when the environment changes, or under conditions of increasing volatility and uncertainty (Dayan, Kakade, and Montague 2000; Daw, Niv, and Dayan 2005; Daw et al. 2006; 2011; Kennerley et al. 2006; Behrens et al. 2007; Meder et al. 2017; Massi, Donahue, and Lee 2018). In a stable environment, long timescales of integration increase the accuracy

of estimating the underlying reward probabilities, improving overall reward maximization (Bernacchia et al. 2011; Iigaya et al. 2019). In a compromise between the two extremes, there is evidence that animals use a strategy that relies on a combination of recent and distant history that diverges from traditional RL model predictions (Corrado et al. 2005; Lau and Glimcher 2005; Sugrue 2004; Iigaya et al. 2019).

Overweighting distant history relative to recent history has also been described as the phenomenon of 'undermatching', and has previously been considered a suboptimal form of decision-making in laboratory behaviors (Sugrue 2004; Corrado et al. 2005; Lau and Glimcher 2005). Recent work has examined undermatching as a distinct and quantifiable strategy that provides a bias-variance tradeoff to animals performing behavior tasks in which reward probability varies over time (Iigaya et al. 2019), and found that animals use at least three underlying time-constants to make a decision in their task. Given the wide variety of timescales an animal must navigate in its life, it is intuitive that integration of temporal information varies over a wide range.

One intriguing possibility is that history information integrated across different timescales is encoded in a distributed and heterogenous manner, creating a reservoir of temporal information available to the animal (Bernacchia et al. 2011). Heterogenous temporal encoding provides a mechanism by which an animal may maintain information efficiently with a simple update rule mediated networks of neurons (Fusi, Drew, and Abbott 2005; Soltani 2006; Tiganj, Hasselmo, and Howard 2015), but also maintain independent information about a variety of timescales at once (Bernacchia et al. 2011; Spitmaan et al. 2020). Previous studies have showed that animals encode history-based value related information in a distributed and diverse manner throughout cortical and subcortical structures

(Sul et al. 2010; 2011; Hattori et al. 2019; Bari et al. 2019; Steinmetz et al. 2019) and it follows that behaviorally-relevant temporal information would be represented in this way as well.

What is optimal for the animal, therefore, is to have a flexible neural encoding system that can be both responsive to recent experience but remain sensitive to experience that is more distant in time. Such a neural substrate ought to have encoding for temporal information across a range of timescales, maintain this information in a form that is easily read-out by downstream areas, and be highly interconnected with the decision-making networks in the brain. This dissertation proposes that the retroplenial cortex, RSC, is a strong candidate for such an area.

1.4 Locating history-based value coding in the brain

In value-based decision-making, an animal makes a choice on the basis of an internal representation of value, not an external sensory cue or percept, and therefore the constituent elements in reaching the decision are not directly measurable. When studying the neural coding of value-based decision-making, mathematical modeling of behavior is used to estimate the value functions and the latent variables like value and RPE, as well as controlling parameters like learning rate and sensitivity to value (Lee, Seo, and Jung 2012; Rangel, Camerer, and Montague 2008). Decomposing a decision into its constituent computed variables permits a nuanced screening of decision-related activity in different brain regions (Rangel, Camerer, and Montague 2008), and so the selection of both the behavior task and the model are critical to contextualizing the neural results.

A simple but powerful family of models for how to accumulate experience into a decision are the Q-learning models, or value-learning models, in which an explicitly defined

value is updated in an iterative process (Sutton and Barto 1998; Barraclough, Conroy, and Lee 2004; Samejima et al. 2005). One formalization of this is the Rescorla-Wagner model (Rescorla and Wagner 1972), which has its history in classical and operant conditioning—specifically, in the association of an action to a reward.

In this model, a reinforcement learning agent learns the expected value of an action, compares the values associated with the actions, then updates the value of the chosen action by the difference between the expected value and the outcome (Figure 1.1). This difference, RPE, is then used to update the previous expectation, scaled by some learning rate.

In the Rescorla-Wagner formulation, the expected value for each option at a given timepoint is updated at each timestep, generally a trial, which makes this a tractable system for laboratory-based experiments that use this structure. One such class of tasks that is often used in conjunction with Q-learning models in general, and Rescorla-Wagner models in particular, are the multi-armed bandit tasks (Stephens and Krebs 1986; Sutton and Barto 1998). In this type of task there are a fixed number of choices which have different underlying probabilities of reward, and the learner performing the bandit task has no certainty about which has a higher payout and relies on trial and error.

The binarized version of this task, a two-armed bandit, includes two choices with different probabilities of reward. When these probabilities change over time, in some cases drifting slowly and in other cases inverting in a block structure, this is referred to as a dynamic foraging task (Sugrue 2004; Samejima et al. 2005; Sul et al. 2010; 2011; Kawai et al. 2015; Hamid et al. 2016; Tsutsui et al. 2016). The probabilities may either move together or independently, depending on the task, and versions of this task have been used for humans (Behrens et al. 2007; Kolling et al. 2012), non-human primates (Sugrue 2004; Samejima et al.

2005; Tsutsui et al. 2016; Kawai et al. 2015), and rodents (Sul et al. 2010; 2011; Hamid et al. 2016; Hattori et al. 2019; Bari et al. 2019), in addition to being a testbed for artificial intelligence research (Kaelbling 1993; Averbeck 2015). Modeling foraging task behavior with a Rescorla-Wagner model provides trial-by-trial readout of the estimated value the animal is using, the difference in value between conditions, as well as information about the decision parameters learning rate and sensitivity to value.

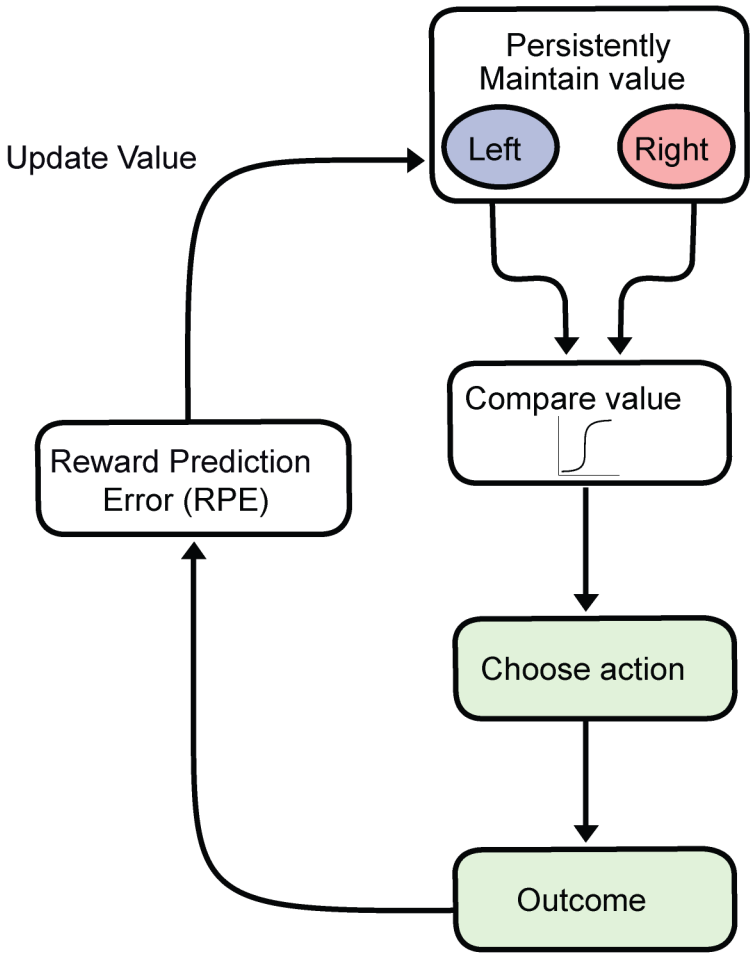


Figure 1.1: Schematic of Rescorla-Wagner Q-learning model for a two-choice case.
Value is updated based on experience, the values associated with two options are compared, an action is selected, the outcome is observed, and the difference between observed and expected is used to update the estimate of value.

1.5 Value-related signals are widespread and heterogeneously encoded

Given an estimate of the latent variables of decision-making (i.e. value and RPE), researchers can then identify correlates of these variables in the brain. There is broad evidence that multiple areas of the brain encode various aspects of decision-making, and that some areas may encode multiple steps of the process (Lee, Seo, and Jung 2012). Dissecting these findings provides a roadmap for where and how decision is made at the level of neural populations.

The estimated value of the chosen option is central to RL theories of behavior, and expected to both represent the integration of past choices and predict the upcoming choices an animal will make. Activity related to the chosen value has been observed in dorsolateral prefrontal cortex in humans and primates (Daw et al. 2006; Barraclough, Conroy, and Lee 2004; So and Stuphorn 2010), medial frontal cortex in primates and rodents (Seo and Lee 2009; Sul et al. 2010; 2011), parietal cortex in primates (Platt and Glimcher 1999; Sugrue 2004; Dorris and Glimcher 2004), supplementary or pre-motor cortex in primates and rodents (Sul et al. 2011; Pastor-Bernier and Cisek 2011; Bari et al. 2019), as well as the striatum across all model organisms (Samejima et al. 2005; Daw et al. 2006; Lau and Glimcher 2008; H. Kim et al. 2009; M. Ito and Doya 2009; Cai, Kim, and Lee 2011).

A second critical component of RL theory is the update rule, governed in these models by RPE. It is expected that brain areas involved with updating estimates of value would encode both the chosen value and the RPE, a function of the value. Estimated value would be expected to vary gradually over time according to this updating signal from RPE. Unlike estimated value, which is a signal that should be relatively persistent, RPE instead ought to be transient and might only be experienced immediately following the evaluation of reward. Transient signals correlated with RPE were first identified in midbrain dopamine neurons

(Schultz 2000), and have since been largely associated with dopamine in wide-ranging projections through the striatum (Bayer and Glimcher 2005; H. Kim et al. 2009; Li and Daw 2011; Hamid et al. 2016; Dabney et al. 2020; H. R. Kim et al. 2020) and cortex (Padoa-Schioppa and Assad 2006; Matsumoto et al. 2007; Seo and Lee 2007; Sul et al. 2010; Dabney et al. 2020). Updating the expected value would likely be mediated by an area in which signals related to the chosen value and the reward prediction error converge (H. Kim et al. 2009; Sul et al. 2010; Tsutsui et al. 2016).

In addition, functions of value—such as the direct comparison between the value of two options, the value difference—strongly identify certain areas as being proximal to the decision (Lee, Seo, and Jung 2012). In RL models the probability of choosing between two options is determined by their value difference, and so neurons that encode value difference may be downstream of value update and maintenance, but upstream of motor output (Sul et al. 2011; Bari et al. 2019; Hattori et al. 2019). Perturbational experiments that disrupted the activity of neurons in the pre-motor cortex (Sul et al. 2011), medial prefrontal cortex (Bari et al. 2019), and retrosplenial cortex (Hattori et al. 2019) of rodents all quantifiably impaired the use of value information in decision-making.

Guided by the encoding of latent variables of value, targeted perturbational experiments can be used to dissect how history-based value information is compared, maintained, and used in the brain.

1.6 Area-specificity and plasticity of history-dependent value coding during learning

Foundational to this dissertation work was Hattori et al. 2019, which surveyed the differences in value coding across six cortical areas, described how these differences emerged

with learning, and implicated retrosplenial cortex (RSC) as a critical node in the decision-making network. What follows is a brief summary of the relevant findings.

In head-fixed mice performing the dynamic foraging task (Sugrue 2004; Samejima et al. 2005; Sul et al. 2010; 2011; Kawai et al. 2015; Hamid et al. 2016; Tsutsui et al. 2016), neural activity from six dorsal cortical areas was recorded with two-photon calcium imaging, in excitatory cells in layers 2/3 of cortex. The areas of interest were: the pre-motor areas anterolateral motor cortex (ALM), and posterior secondary motor cortex (pM2); the associational areas posterior parietal cortex (PPC), and retrosplenial cortex (RSC); and the sensory areas primary somatosensory cortex (S1), and primary visual cortex (V1).

Adapting a Rescorla-Wagner model specifically for the task (Rescorla and Wagner 1972; Barraclough, Conroy, and Lee 2004; Makoto Ito and Doya 2011) yielded a behavioral model that well-described the mouse behavior, and provided the estimated value associated with two options. This value updated on a trial-by-trial basis in a recursive manner, and was a time-varying signal that could in turn be correlated with the neural activity to identify value-coding neurons. In this binary choice task, there were two sets of time-varying values, value for the left side and value for the right side. This was then further codified as the difference in value, which is the determinant of which side is estimated to have a higher probability of reward, and the value of the chosen side, which indicates whether the mouse is exploiting a higher probability choice or exploring a lower probability choice.

Neurons in all six cortical areas encoded chosen value and value difference, especially in the time window immediately before and following choice. RSC and PPC in particular strongly encoded the chosen value and value difference immediately following choice. Interestingly, in RSC this encoding was particularly persistent and remained high through the

inter-trial interval. In addition, the population-level decoding accuracy of chosen value and value difference was higher for RSC than for any other area throughout the trial, and this population code was consistent throughout the trial.

This value-coding was heterogenous across neurons even within the same area, with some cells encoding value difference but not chosen value, or vice versa, or being positively modulated or negatively modulated to these terms. Some cells also encoded other variables associated with the task like the sum of value, an indication of motivational state, or reward-history, in addition to value difference and chosen value.

Separate from but complimentary to the value coding analysis, population-level decoding of the neural activity revealed that the rewarded-choice history of up to ten past trials was encoded in ALM, PPC, and RSC, but with particular strength in RSC. Modulation by distant history was weakest in the primary sensory areas V1 and S1. When quantified longitudinally across training, the population decoding accuracy of rewarded-choice history information increased in all areas, but most strongly in RSC. Specifically, the population encoding of more distant experience (>3 past trials) was non-existent in the early sessions for all areas, but increased for all cortical areas, and significantly more for RSC than for any other area.

These results from both the single-cell and population level suggested RSC adaptively encoded the specific history information necessary for the value-based decision strategy. While decision variables were widely encoded in cortex, RSC was particularly enriched in cells that encoded value strongly and persistently. This poses RSC as a strong candidate area involved in encoding of history-based value information across timescales.

References

Averbeck, Bruno B. 2015. “Theory of Choice in Bandit, Information Sampling and Foraging Tasks.” Edited by Paul Schrater. *PLOS Computational Biology* 11 (3): e1004164. https://doi.org/10.1371/journal.pcbi.1004164.

Bari, Bilal A., Cooper D. Grossman, Emily E. Lubin, Adithya E. Rajagopalan, Jianna I. Cressy, and Jeremiah Y. Cohen. 2019. “Stable Representations of Decision Variables for Flexible Behavior.” *Neuron* 103 (5): 922-933.e7. https://doi.org/10.1016/j.neuron.2019.06.001.

Barraclough, Dominic J, Michelle L Conroy, and Daeyeol Lee. 2004. “Prefrontal Cortex and Decision Making in a Mixed-Strategy Game.” *Nature Neuroscience* 7 (4): 404–10. https://doi.org/10.1038/nn1209.

Bayer, Hannah M., and Paul W. Glimcher. 2005. “Midbrain Dopamine Neurons Encode a Quantitative Reward Prediction Error Signal.” *Neuron* 47 (1): 129–41. https://doi.org/10.1016/j.neuron.2005.05.020.

Behrens, Timothy E J, Mark W Woolrich, Mark E Walton, and Matthew F S Rushworth. 2007. “Learning the Value of Information in an Uncertain World.” *Nature Neuroscience* 10 (9): 1214–21. https://doi.org/10.1038/nn1954.

Bernacchia, Alberto, Hyojung Seo, Daeyeol Lee, and Xiao-Jing Wang. 2011. “A Reservoir of Time Constants for Memory Traces in Cortical Neurons - Supplement.” *Nature Neuroscience* 14 (3): 366–72. https://doi.org/10.1038/nn.2752.

Cai, Xinying, Soyoun Kim, and Daeyeol Lee. 2011. “Heterogeneous Coding of Temporally Discounted Values in the Dorsal and Ventral Striatum during Intertemporal Choice.” *Neuron* 69 (1): 170–82. https://doi.org/10.1016/j.neuron.2010.11.041.

Corrado, Greg S., Leo P. Sugrue, H. Sebastian Seung, and William T. Newsome. 2005. “Linear-Nonlinear-Poisson Models of Primate Choice Dynamics.” *Journal of the Experimental Analysis of Behavior* 84 (3): 581–617. https://doi.org/10.1901/jeab.2005.23-05.

Dabney, Will, Zeb Kurth-Nelson, Naoshige Uchida, Clara Kwon Starkweather, Demis Hassabis, Rémi Munos, and Matthew Botvinick. 2020. “A Distributional Code for Value in Dopamine-Based Reinforcement Learning.” *Nature* 577 (7792): 671–75. https://doi.org/10.1038/s41586-019-1924-6.

Daw, Nathaniel D, Yael Niv, and Peter Dayan. 2005. “Uncertainty-Based Competition between Prefrontal and Dorsolateral Striatal Systems for Behavioral Control.” *Nature Neuroscience* 8 (12): 1704–11. https://doi.org/10.1038/nn1560.

Daw, Nathaniel D., John P. O’Doherty, Peter Dayan, Ben Seymour, and Raymond J. Dolan. 2006. “Cortical Substrates for Exploratory Decisions in Humans.” *Nature* 441 (7095): 876–79. https://doi.org/10.1038/nature04766.

Daw, Nathaniel D., Samuel J. Gershman, Ben Seymour, Peter Dayan, and Raymond J. Dolan. 2011. “Model-Based Influences on Humans’ Choices and Striatal Prediction Errors.” *Neuron* 69 (6): 1204–15. https://doi.org/10.1016/j.neuron.2011.02.027.

Dayan, Peter, Sham Kakade, and P. Read Montague. 2000. “Learning and Selective Attention.” *Nature Neuroscience* 3 (S11): 1218–23. https://doi.org/10.1038/81504.

Dorris, Michael C., and Paul W. Glimcher. 2004. “Activity in Posterior Parietal Cortex Is Correlated with the Relative Subjective Desirability of Action.” *Neuron* 44 (2): 365–78. https://doi.org/10.1016/j.neuron.2004.09.009.

Fusi, Stefano, Patrick J. Drew, and L.F. Abbott. 2005. “Cascade Models of Synaptically Stored Memories.” *Neuron* 45 (4): 599–611. https://doi.org/10.1016/j.neuron.2005.02.001.

Hamid, Arif A, Jeffrey R Pettibone, Omar S Mabrouk, Vaughn L Hetrick, Robert Schmidt, Caitlin M Vander Weele, Robert T Kennedy, Brandon J Aragona, and Joshua D Berke. 2016. “Mesolimbic Dopamine Signals the Value of Work.” *Nature Neuroscience* 19 (1): 117–26. https://doi.org/10.1038/nn.4173.

Hattori, Ryoma, Bethanny Danskin, Zeljana Babic, Nicole Mlynaryk, and Takaki Komiyama. 2019. “Area-Specificity and Plasticity of History-Dependent Value Coding During Learning.” *Cell* 177 (7): 1858-1872.e15. https://doi.org/10.1016/j.cell.2019.04.027.

Iigaya, Kiyohito, Yashar Ahmadian, Leo P. Sugrue, Greg S. Corrado, Yonatan Loewenstein, William T. Newsome, and Stefano Fusi. 2019. “Deviation from the Matching Law Reflects an Optimal Strategy Involving Learning over Multiple Timescales.” *Nature Communications* 10 (1): 1466. https://doi.org/10.1038/s41467-019-09388-3.

Ito, M., and K. Doya. 2009. “Validation of Decision-Making Models and Analysis of Decision Variables in the Rat Basal Ganglia.” *Journal of Neuroscience* 29 (31): 9861–74. https://doi.org/10.1523/JNEUROSCI.6157-08.2009.

Ito, Makoto, and Kenji Doya. 2011. “Multiple Representations and Algorithms for Reinforcement Learning in the Cortico-Basal Ganglia Circuit.” *Current Opinion in Neurobiology* 21 (3): 368–73. https://doi.org/10.1016/j.conb.2011.04.001.

Kaelbling, Leslie Pack. 1993. *Learning in Embedded Systems*. Cambridge, MA: MIT Press.

Kawai, Takashi, Hiroshi Yamada, Nobuya Sato, Masahiko Takada, and Masayuki Matsumoto. 2015. “Roles of the Lateral Habenula and Anterior Cingulate Cortex in Negative Outcome Monitoring and Behavioral Adjustment in Nonhuman Primates.” *Neuron* 88 (4): 792–804. https://doi.org/10.1016/j.neuron.2015.09.030.

Kennerley, Steven W, Mark E Walton, Timothy E J Behrens, Mark J Buckley, and Matthew F S Rushworth. 2006. "Optimal Decision Making and the Anterior Cingulate Cortex." *Nature Neuroscience* 9 (7): 940–47. https://doi.org/10.1038/nn1724.

Kim, H., J. H. Sul, N. Huh, D. Lee, and M. W. Jung. 2009. "Role of Striatum in Updating Values of Chosen Actions." *Journal of Neuroscience* 29 (47): 14701–12. https://doi.org/10.1523/JNEUROSCI.2728-09.2009.

Kim, HyungGoo R., Athar N. Malik, John G. Mikhael, Pol Bech, Iku Tsutsui-Kimura, Fangmiao Sun, Yajun Zhang, et al. 2020. "A Unified Framework for Dopamine Signals across Timescales." *Cell* 183 (6): 1600-1616.e25. https://doi.org/10.1016/j.cell.2020.11.013.

Kolling, Nils, Timothy E. J. Behrens, Rogier B. Mars, and Matthew F. S. Rushworth. 2012. "Neural Mechanisms of Foraging." *Science* 336 (6077): 95–98. https://doi.org/10.1126/science.1216930.

Lau, Brian, and Paul W. Glimcher. 2005. "Dynamic Response-by-response Models of Matching Behavior in Rhesus Monkeys." *Journal of the Experimental Analysis of Behavior* 84 (3): 555–79. https://doi.org/10.1901/jeab.2005.110-04.

Lau, Brian, and Paul W. Glimcher. 2008. "Value Representations in the Primate Striatum during Matching Behavior." *Neuron* 58 (3): 451–63. https://doi.org/10.1016/j.neuron.2008.02.021.

Lee, Daeyeol, Hyojung Seo, and Min Whan Jung. 2012. "Neural Basis of Reinforcement Learning and Decision Making." *Annual Review of Neuroscience* 35 (1): 287–308. https://doi.org/10.1146/annurev-neuro-062111-150512.

Li, J., and N. D. Daw. 2011. "Signals in Human Striatum Are Appropriate for Policy Update Rather than Value Prediction." *Journal of Neuroscience* 31 (14): 5504–11. https://doi.org/10.1523/JNEUROSCI.6316-10.2011.

Massi, Bart, Christopher H. Donahue, and Daeyeol Lee. 2018. "Volatility Facilitates Value Updating in the Prefrontal Cortex." *Neuron* 99 (3): 598-608.e4. https://doi.org/10.1016/j.neuron.2018.06.033.

Matsumoto, Madoka, Kenji Matsumoto, Hiroshi Abe, and Keiji Tanaka. 2007. "Medial Prefrontal Cell Activity Signaling Prediction Errors of Action Values." *Nature Neuroscience* 10 (5): 647–56. https://doi.org/10.1038/nn1890.

Meder, David, Nils Kolling, Lennart Verhagen, Marco K. Wittmann, Jacqueline Scholl, Kristoffer H. Madsen, Oliver J. Hulme, Timothy E.J. Behrens, and Matthew F.S. Rushworth. 2017. "Simultaneous Representation of a Spectrum of Dynamically Changing Value Estimates during Decision Making." *Nature Communications* 8 (1): 1942. https://doi.org/10.1038/s41467-017-02169-w.

Padoa-Schioppa, Camillo, and John A. Assad. 2006. “Neurons in the Orbitofrontal Cortex Encode Economic Value.” *Nature* 441 (7090): 223–26. https://doi.org/10.1038/nature04676.

Pastor-Bernier, A., and P. Cisek. 2011. “Neural Correlates of Biased Competition in Premotor Cortex.” *Journal of Neuroscience* 31 (19): 7083–88. https://doi.org/10.1523/JNEUROSCI.5681-10.2011.

Platt, Michael L., and Paul W. Glimcher. 1999. “Neural Correlates of Decision Variables in Parietal Cortex.” *Nature* 400 (6741): 233–38. https://doi.org/10.1038/22268.

Rangel, Antonio, Colin Camerer, and P. Read Montague. 2008. “A Framework for Studying the Neurobiology of Value-Based Decision Making.” *Nature Reviews Neuroscience* 9 (7): 545–56. https://doi.org/10.1038/nrn2357.

Rescorla, R., and A. Wagner. 1972. “A Theory of Pavlovian Conditioning : Variations in the Effectiveness of Reinforcement and Nonreinforcement.” In *Classical Conditioning II: Current Research and Theory*, 64–99. Appleton-Century-Crofts.

Samejima, Kazuyuki, Yasumasa Ueda, Kenji Doya, and Minoru Kimura. 2005. “Representation of Action-Specific Reward Values in the Striatum.” *Science* 310 (5752): 1337–40. https://doi.org/10.1126/science.1115270.

Schultz, W. 2000. “Reward Processing in Primate Orbitofrontal Cortex and Basal Ganglia.” *Cerebral Cortex* 10 (3): 272–83. https://doi.org/10.1093/cercor/10.3.272.

Seo, H., and D. Lee. 2007. “Temporal Filtering of Reward Signals in the Dorsal Anterior Cingulate Cortex during a Mixed-Strategy Game.” *Journal of Neuroscience* 27 (31): 8366–77. https://doi.org/10.1523/JNEUROSCI.2369-07.2007.

Seo, H., and D. 2009. “Behavioral and Neural Changes after Gains and Losses of Conditioned Reinforcers.” *Journal of Neuroscience* 29 (11): 3627–41. https://doi.org/10.1523/JNEUROSCI.4726-08.2009.

So, Na-Young, and Veit Stuphorn. 2010. “Supplementary Eye Field Encodes Option and Action Value for Saccades With Variable Reward.” *Journal of Neurophysiology* 104 (5): 2634–53. https://doi.org/10.1152/jn.00430.2010.

Soltani, A. 2006. “A Biophysically Based Neural Model of Matching Law Behavior: Melioration by Stochastic Synapses.” *Journal of Neuroscience* 26 (14): 3731–44. https://doi.org/10.1523/JNEUROSCI.5159-05.2006.

Spitmaan, Mehran, Hyojung Seo, Daeyeol Lee, and Alireza Soltani. 2020. “Multiple Timescales of Neural Dynamics and Integration of Task-Relevant Signals across Cortex.” *Proceedings of the National Academy of Sciences* 117 (36): 22522–31. https://doi.org/10.1073/pnas.2005993117.

Steinmetz, Nicholas A., Peter Zatka-Haas, Matteo Carandini, and Kenneth D. Harris. 2019. "Distributed Coding of Choice, Action and Engagement across the Mouse Brain." *Nature* 576 (7786): 266–73. https://doi.org/10.1038/s41586-019-1787-x.

Stephens, D. W., and J. R. Krebs. 1986. *Foraging Theory*. Princeton, NJ: Princeton University Press.

Sugrue, L. P. 2004. "Matching Behavior and the Representation of Value in the Parietal Cortex." *Science* 304 (5678): 1782–87. https://doi.org/10.1126/science.1094765.

Sul, Jung Hoon, Suhyun Jo, Daeyeol Lee, and Min Whan Jung. 2011. "Role of Rodent Secondary Motor Cortex in Value-Based Action Selection." *Nature Neuroscience* 14 (9): 1202–8. https://doi.org/10.1038/nn.2881.

Sul, Jung Hoon, Hoseok Kim, Namjung Huh, Daeyeol Lee, and Min Whan Jung. 2010. "Distinct Roles of Rodent Orbitofrontal and Medial Prefrontal Cortex in Decision Making." *Neuron* 66 (3): 449–60. https://doi.org/10.1016/j.neuron.2010.03.033.

Sutton, Richard S., and Andrew G. Barto. 1998. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Tiganj, Zoran, Michael E. Hasselmo, and Marc W. Howard. 2015. "A Simple Biophysically Plausible Model for Long Time Constants in Single Neurons: A SIMPLE BIOPHYSICALLY PLAUSIBLE MODEL FOR LONG TIME CONSTANTS." *Hippocampus* 25 (1): 27–37. https://doi.org/10.1002/hipo.22347.

Tsutsui, Ken-Ichiro, Fabian Grabenhorst, Shunsuke Kobayashi, and Wolfram Schultz. 2016. "A Dynamic Code for Economic Object Valuation in Prefrontal Cortex Neurons." *Nature Communications* 7 (1): 12554. https://doi.org/10.1038/ncomms12554.

Yu, Angela J., and Peter Dayan. 2005. "Uncertainty, Neuromodulation, and Attention." *Neuron* 46 (4): 681–92. https://doi.org/10.1016/j.neuron.2005.04.026.

Following from the result that history-based value was widely encoded in cortex but RSC showed unique strength and persistency of value-coding (Hattori et al. 2019), we speculated that RSC might have a privileged role in the decision-making circuit, and therefore that the behavior would be sensitive to the loss of neural activity in RSC. To test whether RSC was necessary for task performance, we performed optogenetic inactivation of RSC via light-targeted activation of parvalbumin-positive inhibitory interneurons in PV-Cre::LSL-ChR2 transgenic mice (Figure 2.1 A). Light illuminated the cortical surface in specific patterns directed by a projector-based system (Dhawale et al. 2010; Haddad et al. 2013), which allowed specific inactivation of RSC along its full rostro-caudal axis (Figure 2.1 B).

2.2 Results

We found that RSC inactivation decreased both the probability of repeating the same action after rewarded trials (‘‘win-stay’’) and the probability of changing action after unrewarded trials (‘‘lose-switch’’) (Figure 2.1 C), indicating the animal had an impaired association between its immediately preceding choices and rewards.

Logistic regression analysis showed that the inactivation also attenuated the mouse’s reliance on rewarded- and unrewarded-choice history (Figure 2.1 D, E), but had no effect on perseverative choice (Figure 2.1 F) or choice bias. These results suggested that neural activity in RSC was necessary for the reward-history-dependent decision-making strategy.

Given the observation from the imaging results that history- and value-related signals were widespread across cortex, we tested the possibility that other areas might compensate for the function of RSC when it is fully removed. We performed chronic, bilateral lesions of RSC by injecting N-Methyl-D-aspartate (NMDA) to induce excitotoxicity (Figure 2.1 G). Unlike

the acute inactivation with optogenetics, the RSC lesion at expert stage did not affect the behavioral performance in subsequent sessions. The win-stay probabilities were unchanged following lesion (Figures 2.1 H and 2.2 A), and the change in logistic weights following lesion were no different from the saline-injected sham animals (Figures 2.1 I and 2.2 B). Neither metric showed any relation to lesion size (Figure 2.2 A, C). Remarkably, compensation occurred from even the very first day following lesion (Figure 2.2 D, E), with no re-learning required.

These results indicate that RSC is acutely necessary for the decision-making task, but that the interconnected nature of the decision-making system and widespread encoding of value means that in the chronic absence of RSC other areas might compensate. Given the ethological salience of decision-making, it is not surprising that this is a conserved, widely distributed, and robust system. Even still, inactivating RSC during decision-making selectively impaired the reward-history-based strategy, suggesting RSC is involved when the animal incorporates and uses experience from past trials.

This study revealed RSC as a critical region for decision making based on history-dependent value, and indicated it is a rich and heterogenous cortical area for further study.

Figure 2.1: Acute inactivation of RSC, but not its chronic lesion, impairs reward history-based strategy

(A) Schematic of the projector-based optical stimulation system. Patterned light is resized and focused on cortex to optogenetically activate parvalbumin-positive inhibitory neurons.

(B) RSC was bilaterally inactivated in a small subset of trials within a session (5% or 15% of trials). In all other trials, the head bar was illuminated with the same light intensity and area. Elliptic illumination patterns were used for RSC inactivation trials to cover rostro-caudally elongated RSC. The illumination was applied from the onset of ready period until the choice at 30 Hz with a linear attenuation in the intensity after choice.

(C) Effects of RSC inactivation on the win-stay and lose-switch probabilities (left: n = 5 mice; right: n = 12 sessions). Red line indicates the mean of each condition. Only successive choice trials were used to derive the probabilities. P(Win-stay) was normalized by the overall stay probability (the average of P(Win-stay) and P(Lose-stay)). Similarly, P(Lose-switch) was normalized by the overall switch probability (the average of P(Win-switch) and P(Lose-switch)). For the n = animals plots (left), all sessions from each mouse were pooled to calculate the probabilities. For the n = sessions plots (right), only pairs from the 15% inactivation sessions were included. RSC inactivation made the stay and switch probabilities less dependent on the reward outcomes from the 1 trials. Paired t test.

(D) Behavioral dependency on rewarded choice (RewC(t-i)), unrewarded choice (UnrC(t-i)), and outcome-independent choice (C(t-i)) history in head bar trials and RSC inactivation trials (STAR Methods, Equation 23).

(E) Effects of RSC inactivation on behavioral dependency on the 3 types of history from 1 trial (left: n = 5 mice; right: n = 15 sessions). Pairs of head bar trials (black) and RSC inactivation trials (blue) are shown. Red lines indicate the means. RSC inactivation reduced behavioral dependency on choice-reward history, especially for the rewarded choice history. Wilcoxon signed-rank test was used for non-normally distributed UnrC(t-1) weights of n = sessions, and paired t test was used for the other comparisons.

(F) Effect of RSC inactivation to choice bias (left: n = 5 mice; right: n = 15 sessions). The absolute value of bias is shown for pairs of head bar trials (black) and RSC inactivation trials (blue). Red line indicates the mean of each condition. Paired t test. The sign of the bias was also generally unaffected by inactivation (not shown).

(G) (Top) Example coronal section from a mouse with lesioned RSC. The section is stained with NeuN to visualize the presence of neurons. RSC largely lacks NeuN-positive neurons. Dashed lines indicate the borders of RSC. (Bottom) A corresponding brain atlas is shown. Yellow lines outline RSC. Purple shading indicates lesioned area.

(H) Effects of RSC lesion on win-stay and lose-switch probabilities (sham: n = 5 mice; lesion: n = 6 mice). Difference between the mean of 7 sessions before sham or lesion and the mean of 7 sessions after sham or lesion is shown. Red lines indicate the means. Two-sided t test.

(I) Effects of RSC lesion on behavioral dependency on the 3 types of history from 1 trial. Two-sided t test. $p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

(Hattori et al., 2019)

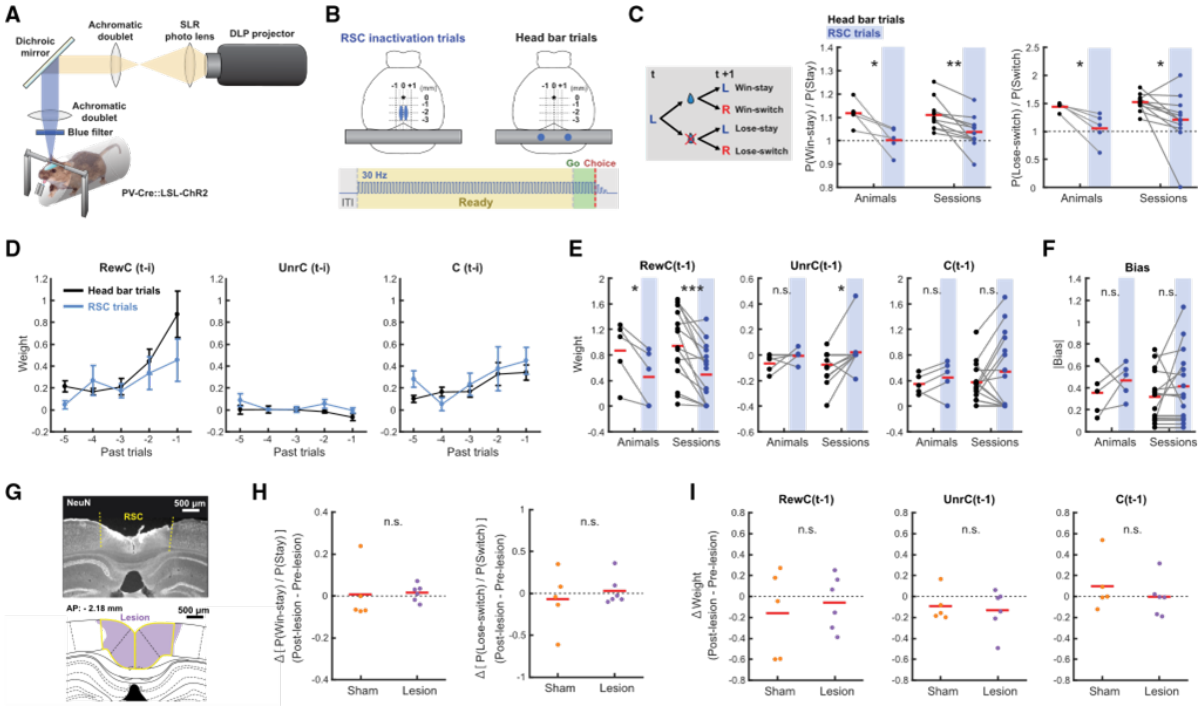

A
Dichroic mirror
Achromatic doublet
SLR photo lens
DLP projector
Achromatic doublet
Blue filter
PV-Cre::LSL-ChR2
B
RSC inactivation trials
Head bar trials
30 Hz
ITI
Ready
Go
Choice
C
Head bar trials
RSC trials
t
t +1
L
Win-stay
R
Win-switch
L
Lose-stay
R
Lose-switch
P(Win-stay) / P(Stay)
P(Lose-switch) / P(Switch)
Animals
Sessions
D
RewC (t-i)
UnrC (t-i)
C (t-i)
Head bar trials
RSC trials
Weight
Past trials
E
RewC(t-1)
UnrC(t-1)
C(t-1)
Weight
Animals
Sessions
F
Bias
|Bias|
Animals
Sessions
G
NeuN
RSC
500 µm
AP: - 2.18 mm
Lesion
500 µm
H
Δ [P(Win-stay) / P(Stay)]
(Post-lesion - Pre-lesion)
Δ [P(Lose-switch) / P(Switch)]
(Post-lesion - Pre-lesion)
Sham
Lesion
I
RewC(t-1)
UnrC(t-1)
C(t-1)
Δ Weight
(Post-lesion - Pre-lesion)
Sham
Lesion

Figure 2.2: RSC Lesion Does Not Impair Reward History-Based Strategy

(A) Relationship between RSC lesion size and the effect on win-stay and lose-switch probabilities. Difference between the mean of 7 sessions before lesion and the mean of 7 sessions after lesion is shown. Pearson's correlation coefficients and their p values are shown.

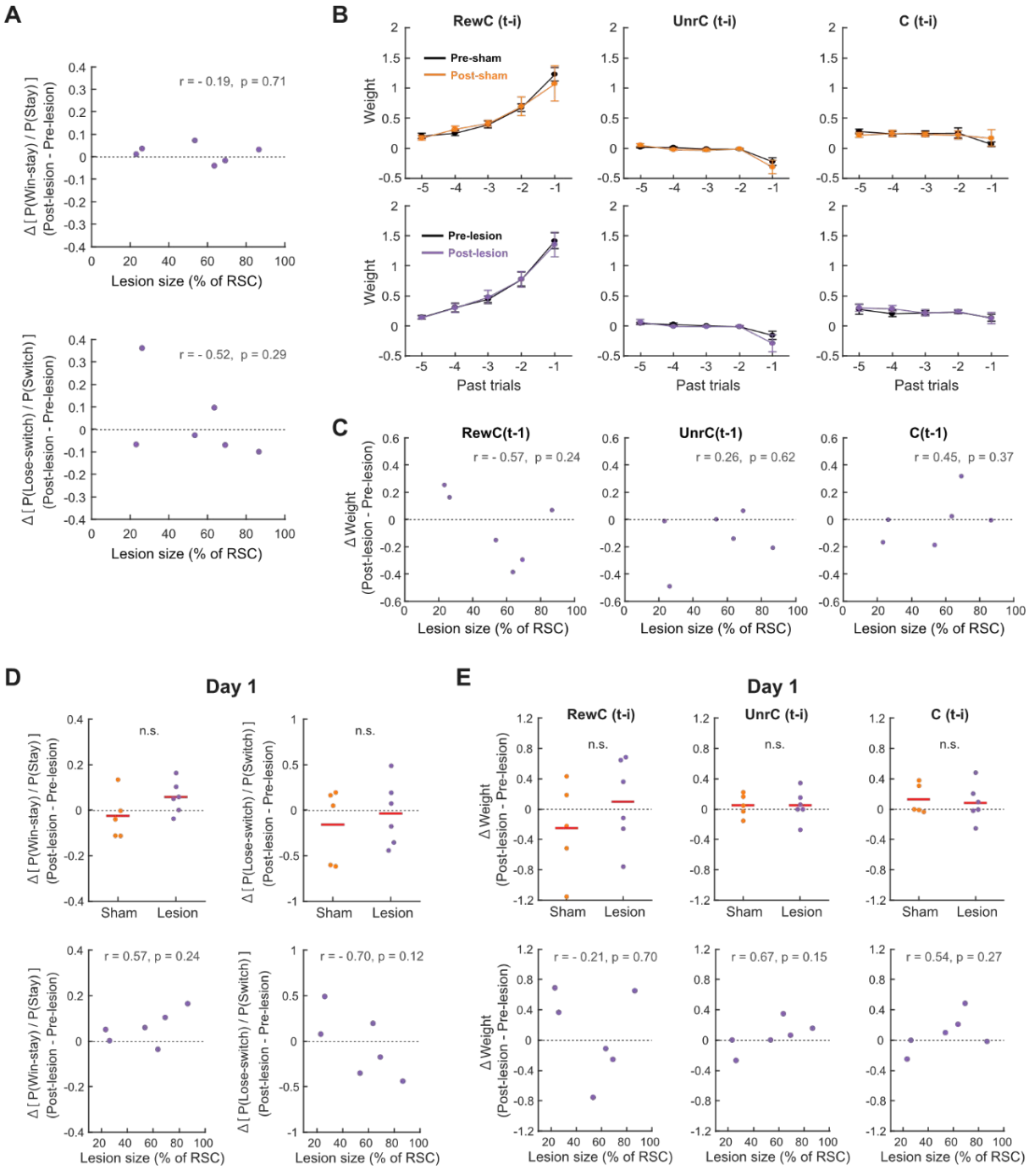
(B) Behavioral dependency on rewarded choice (RewC(t-i)), unrewarded choice (UnrC(t-i)), and outcome-independent choice (C(t-i)) history before and after sham or lesion surgery. The mean of 7 sessions before sham or lesion and the mean of 7 sessions after sham or lesion are shown. The means were averaged across mice (n = 5 sham mice; n = 6 lesion mice). Error bars are SEM.

(C) Relationship between RSC lesion size and the effects on behavioral dependency on the 3 types of history from 1 trial. Difference between the mean of 7 sessions before lesion and the mean of 7 sessions after lesion is shown. Pearson's correlation coefficients and their p values are shown.

(D) Effects of RSC lesion on win-stay and lose-switch probabilities on day 1 after lesion. (Top) Difference between the mean of 7 sessions before sham or lesion and day 1 after sham or lesion is shown (Two-sided t test). Red lines indicate the means. (Bottom) Relationship between RSC lesion size and the effects on win-stay and lose-switch probabilities. Difference between the mean of 7 sessions before lesion and day 1 after lesion is shown. Pearson's correlation coefficients and their p values are shown.

(E) Effects of RSC lesion on behavioral dependency on the 3 types of history from 1 trial. (Top) Difference between the mean of 7 sessions before sham or lesion and day 1 after sham or lesion is shown (Two-sided t test). Red lines indicate the means. (Bottom) Relationship between RSC lesion size and the effects on behavioral dependency on the 3 types of history from 1 trial. Pearson's correlation coefficients and their p values are shown. n.s. $p > 0.05$.

(Hattori et al., 2019)


A
r = - 0.19, p = 0.71
Δ [P(Win-stay) / P(Stay)]
(Post-lesion - Pre-lesion)
Lesion size (% of RSC)
r = - 0.52, p = 0.29
Δ [P(Lose-switch) / P(Switch)]
(Post-lesion - Pre-lesion)
Lesion size (% of RSC)
B
RewC (t-i)
UnrC (t-i)
C (t-i)
Pre-sham
Post-sham
Pre-lesion
Post-lesion
Weight
Past trials
C
RewC(t-1)
UnrC(t-1)
C(t-1)
r = - 0.57, p = 0.24
r = 0.26, p = 0.62
r = 0.45, p = 0.37
Δ Weight
(Post-lesion - Pre-lesion)
Lesion size (% of RSC)
D
Day 1
n.s.
Δ [P(Win-stay) / P(Stay)]
(Post-lesion - Pre-lesion)
Δ [P(Lose-switch) / P(Switch)]
(Post-lesion - Pre-lesion)
Sham
Lesion
r = 0.57, p = 0.24
r = - 0.70, p = 0.12
Lesion size (% of RSC)
E
Day 1
RewC (t-i)
UnrC (t-i)
C (t-i)
n.s.
Δ Weight
(Post-lesion - Pre-lesion)
Sham
Lesion
r = - 0.21, p = 0.70
r = 0.67, p = 0.15
r = 0.54, p = 0.27
Lesion size (% of RSC)

2.5 Selected methods

Excerpted methods from Hattori et al., 2019, relevant to Figures 2.1 and 2.2.

2.5.1 Optogenetic inactivation

To activate PV-positive inhibitory neurons in RSC of PV-Cre::LSL-ChR2 double transgenic mice using optogenetics, we generated elliptical illumination patterns with a DLP projector (Optoma X600 XGA). A single-lens reflex (SLR) lens (Nikon, 50 mm, f/1.4D, AF) was coupled with 2 achromatic doublets (Thorlabs, AC508-150-A-ML, f = 150 mm; Thorlabs, AC508-075-A-ML, f = 75 mm) to shrink and focus illumination patterns on RSC. A dichroic mirror (Thorlabs, DMLP490L) and a blue filter (Thorlabs, FESH0450) were placed between the 2 achromatic doublets and after the 2nd achromatic doublet, respectively, to pass only blue light (400-450 nm). Illumination patterns were generated with Psychtoolbox in MATLAB (http://psychtoolbox.org/). In RSC inactivation trials, a 2 mm 3 0.5 mm ellipse was focused on RSC in each hemisphere (Center at 0.3 mm lateral and 2 mm posterior to bregma). In all other trials, two 1 mm 3 1 mm circles were focused on the head bar ('head bar trials'). The total light intensity was equivalent between RSC inactivation trials and head bar trials. We projected the patterns at 30 Hz as a sequence of square pulses from the onset of the ready period until the choice, with a linear attenuation in intensity over the last 100 ms. The intensity at the focus ranged between 2.5-6 mW/mm2 to moderately activate ChR2-expressing neurons (Dhawale et al., 2010; Haddad et al., 2013). We set the frequency of RSC inactivation trials within a session to either 15% (12 sessions) or 5% (3 sessions) with the constraint that each RSC inactivation trial must be followed by at least 3 head bar trials to avoid excessive perturbation of reinforcement learning. We inactivated RSC through a glass window for 4 mice and through the skull for 1 mouse. The skull for the through-skull

inactivation was made semi-transparent by covering the dorsal skull surface with a layer of cyanoacrylate glue (Makino et al., 2017).

2.5.2 Lesion

Twelve adult mice were trained to perform the task, and after at least 7 days of stable performance underwent excitotoxic-lesion or sham or lesion surgery. Stable, expert performance for this task was determined to be choice prediction accuracy of > 65% with a standard RL model (Equation 3 and Equation 4) in at least 6 sessions during the 7 days; these sessions also met the > 0.08 RL index criterion for imaging mice in at least 6 sessions during the 7 days. Mice were anesthetized with 1%–2% isoflurane during surgery. Three burr-hole craniotomies per hemisphere (6 total) were opened on the dorsal skull over RSC. A tapered glass pipette was inserted to perform the cortical microinjection. Injection sites were, in mm and relative to Bregma: AP = 1.6, 2.3, 3.0, ML = ± 0.3, ± 0.35, ± 0.4, and DV = 0.4 from the dura surface in all sites. Injection was of 50 nL/site of either NMDA in sterile saline (20 mg/ml or 10 mg/ml; Sigma) or sterile saline, at a rate of 0.05-0.1 ml/min. After injection, the pipette was left for 5 min to ensure diffusion of the solution. Buprenorphine (0.1 mg/kg of body weight) and Baytril (10 mg/kg of body weight) were subcutaneously injected after surgery. Following surgery, the mouse resumed the behavioral task on the next day, and thereafter every day. Both the surgeon and the experimenter for the behavior were blind to the identity of the substance that was injected, and became unblinded only after the last day of data collection. Of the 12 mice, 5 received saline, 7 received NMDA. One of the NMDA mice was excluded due to small and off-target lesion, as quantified by histology. Brains of lesion and saline mice were collected at 21-25 days post injection. To quantify the lesion size, 50 mm-thick coronal sections were prepared with a microtome (Thermo Fisher Scientific) and

blocked with 10% goat serum, 1% bovine serum albumin, and 0.3% Triton X-100 in PBS. Immunostaining was then performed with anti-NeuN primary antibody (1:400; Mouse, Millipore) and anti-mouse Alexa Fluor 488 secondary antibody (1:1000; Goat, Thermo Fisher Scientific). Both missing areas and areas that lacked NeuN-positive neurons were considered lesioned. Images of coronal sections with RSC and the corresponding brain atlas (Paxinos and Franklin, 2001) were superimposed to quantify the % of lesion within RSC.

2.5.3 Effects of optogenetic RSC inactivation on behavioral history dependency

To quantify the effects of RSC inactivation on the behavioral history dependency, we fit the following logistic regression model:

$$
\begin{aligned}
logit\big(P_L(t)\big) = &\left(\sum_{i=1}^{5} \beta^{HB}_{RewC(t-i)} * RewC(t-i) + \sum_{i=1}^{5} \beta^{HB}_{UnrC(t-i)} * UnrC(t-i)\right. \\
&\left. + \sum_{i=1}^{5} \beta^{HB}_{C(t-i)} * C(t-i) + \beta^{HB}_{0}\right) * HB(t) \\
&+ \left(\sum_{i=1}^{5} \beta^{RSC}_{RewC(t-i)} * RewC(t-i) + \sum_{i=1}^{5} \beta^{RSC}_{UnrC(t-i)} * UnrC(t-i)\right. \\
&\left. + \sum_{i=1}^{5} \beta^{RSC}_{C(t-i)} * C(t-i) + \beta^{RSC}_{0}\right) * RSC(t) \qquad \textbf{[eq. 23]}
\end{aligned}
$$

where $RewC(t-i)$ is the rewarded choice history on trial $t-i$ (1 if rewarded left choice, -1 if rewarded right choice, 0 otherwise), $UnrC(t-i)$ is the unrewarded choice history on trial $t-i$ (1 if unrewarded left choice, -1 if unrewarded right choice, 0 otherwise), $C(t-i)$ is the outcome-independent choice history on trial $t-i$ (1 if left choice, -1 if right choice, 0 otherwise). $HB(t)$ is 1 on head bar trials and 0 on RSC inactivation trials. $RSC(t)$ is 1 on RSC inactivation trials and 0 on head bar trials. $\beta_{RewC(t-i)}$, $\beta_{UnrC(t-i)}$, and $\beta_{C(t-i)}$ are

the regression weights of each history predictor, and β_0 is the history-independent constant bias. The model has separate regression weights for head bar and RSC inactivation trials. The model was regularized with L1-penalty where the regularization parameter was selected by 10-fold cross-validation (minimum cross-validation error). To prevent overpenalization of regression weights for less frequent RSC inactivation trials, we matched the number of head bar trials to the number of RSC inactivation trials for each fitting by randomly subsampling head bar trials. The subsampling and fitting were repeated with the smallest number of iterations to include every head bar trial at least once, and the regression weights from the iterations were averaged.

2.5.4 Effects of RSC lesion to behavioral history dependency

To quantify the effects of RSC lesion to the behavioral history dependency, we fit the following logistic regression model:

$$logit\big(P_L(t)\big) = \sum_{i=1}^{5} \beta_{RewC(t-i)} * RewC(t-i) + \sum_{i=1}^{5} \beta_{UnrC(t-i)} * UnrC(t-i) + \sum_{i=1}^{5} \beta_{C(t-i)} * C(t-i) + \beta_0 \qquad [\mathbf{eq.24}]$$

The model was regularized with L1-penalty where the regularization parameter was selected by 10-fold cross-validation (minimum cross-validation error). The model was fit to the choice patterns of each session.

2.6 Acknowledgements

Chapter 2, in part, contains the reproduction of material as it appears in Hattori R, Danskin B, Babic Z, Mlynaryk N, and Komiyama T (2019). Area-Specificity and Plasticity of History-Dependent Value Coding During Learning. *Cell* 177, 1-15. The dissertation author was an author of this paper, and the primary researcher for the optogenetic and lesion experiments detailed in the included figures and selected methods.

References

Dhawale, Ashesh K, Akari Hagiwara, Upinder S Bhalla, Venkatesh N Murthy, and Dinu F Albeanu. 2010. “Non-Redundant Odor Coding by Sister Mitral Cells Revealed by Light Addressable Glomeruli in the Mouse.” *Nature Neuroscience* 13 (11): 1404–12. https://doi.org/10.1038/nn.2673.

Haddad, Rafi, Anne Lanjuin, Linda Madisen, Hongkui Zeng, Venkatesh N Murthy, and Naoshige Uchida. 2013. “Olfactory Cortical Neurons Read out a Relative Time Code in the Olfactory Bulb.” *Nature Neuroscience* 16 (7): 949–57. https://doi.org/10.1038/nn.3407.

Hattori, Ryoma, Bethanny Danskin, Zeljana Babic, Nicole Mlynaryk, and Takaki Komiyama. 2019. “Area-Specificity and Plasticity of History-Dependent Value Coding During Learning.” *Cell* 177 (7): 1858-1872.e15. https://doi.org/10.1016/j.cell.2019.04.027.

Makino, Hiroshi, Chi Ren, Haixin Liu, An Na Kim, Neehar Kondapaneni, Xin Liu, Duygu Kuzum, and Takaki Komiyama. 2017. “Transformation of Cortex-Wide Emergent Properties during Motor Learning.” *Neuron* 94 (4): 880-890.e8. https://doi.org/10.1016/j.neuron.2017.04.015.

Paxinos, George, and Keith BJ Franklin. 2008. *The Mouse Brain in Stereotaxic Coordinates*. Third.

CHAPTER 3. DIVERSE BEHAVIORAL TIMESCALES ENCODED IN RETROSPLENIAL CORTEX EXPLAIN HYPERBOLIC BEHAVIOR

3.1 Summary

Animals rely on their experience to guide their next choice. In foraging-type tasks guided by history-dependent value, these experiences are integrated such that the weights of past events initially decay quickly over time but show a longer tail than expected by exponential decay, which is better described by a hyperbolic function. Hyperbolic integration affords sensitivity to both recent environmental dynamics and long-term trends, however the mechanism by which the brain implements this hyperbolic integration is unknown. We trained mice on a history-dependent, value-based decision task and found that the mice indeed showed hyperbolic decay on their weighting of past experience. However, the activity of history-encoding cortical neurons showed weighting with exponential decay. In resolving this apparent mismatch, we observed that cortical neurons encode history information heterogeneously across a wide variety of time-constants, with the retrosplenial cortex (RSC) overrepresenting longer time-constants than other areas. A model that combines these diverse timescales can recreate the heavy-tailed, hyperbolic-like behavior. In particular, time-constants of RSC neurons best matched the behavior, and optogenetic inactivation of RSC uniquely reduced the use of history information. These results indicate that behavior-relevant history information is maintained in neurons across multiple timescales in parallel, and suggest RSC is a critical reservoir of this information guiding decision-making.

3.2 Introduction

Integrating information from the past to make a decision in the present is a universal and critical component of animal behavior. For instance, in value-based decision making,

animals establish a subjective value for each available action, based on the reward outcomes of actions taken in the recent past. Reinforcement learning (RL) models provide a simple but powerful framework for how to integrate history information to guide future decisions (Sutton and Barto 1998). RL models such as the Rescorla-Wagner model (Rescorla and Wagner 1972) use the difference between expected rewards and observed rewards, known as reward prediction error, to update the subjective estimate of value. In typical formulations, the value associated with the action is updated by combining the new information (reward prediction error) from the most recent trial with the previous value estimates with a fixed learning rate. This update rule weights the influence of recent outcomes more than outcomes in the distant past. Specifically, a fixed learning rate results in exponential decay in the influence of past outcomes on the present estimate in which the influence decays with a fixed ratio for every unit time. Exponential integration of the past is attractive because of its mechanistic simplicity: the brain would in theory only need to update its subjective value by combining, with a fixed rate, the ongoing value representation with reward prediction error.

However, behavior studies across humans (Serences 2008), non-human primates (Sugrue 2004; Corrado et al. 2005; Lau and Glimcher 2005), and other animal models (Aparicio and Baum 2009; Iigaya et al. 2019) engaged in value-based decision making have observed that animal behavior deviates from exponential integration. Specifically, the integration of past experience generally exhibits a sharp initial drop on recent experience with a heavy-tail on more distant experience, which is better fit by a hyperbolic than exponential function. The adaptive advantage of such hyperbolic integration seems intuitive, as the difference in the environment between 1 minute ago and 2 minutes ago is likely more informative about the current environment than the difference between 1 month ago and 1

month plus a minute ago. Thus it is beneficial to weight the experience from 1 minute ago more than 2 minutes ago but the weighting for a month ago and 1 month plus a minute ago should be nearly equivalent, which is achieved by heavy-tailed hyperbolic decay. Scaling the decay of information differentially across time imparts sensitivity to both recent changes and tendencies that are stable long-term. However, the mechanism by which the brain performs hyperbolic-like integration of history is unknown.

To address this issue, we analyzed history integration of cortical neurons in mice engaged in value-based decision making. We find that behavioral integration of history in these mice is more hyperbolic than exponential, similar to previous behavioral studies. Interestingly, however, history integration of individual cortical neurons is more exponential than hyperbolic. We provide a potential explanation for this apparent discrepancy between behavior and neurons by demonstrating that the time-constants of exponential history integration are heterogeneous across neurons. Weighted averaging of these diverse exponential kernels, especially in the retrosplenial cortex (RSC) that overrepresents distant history information compared to other areas, can approximate hyperbolic-like behavioral integration. Inactivation of RSC, but not of posterior parietal cortex (PPC) or posterior premotor cortex (pM2), impairs the use of history information. We propose that RSC neurons function as a pool of heterogeneous exponential history integrators, and appropriate weighting of these neural populations results in adaptive behavior with hyperbolic history integration.

3.2 Results

3.2.1 Mouse choice pattern is better fit by hyperbolic than exponential integration

To investigate the neural basis of history integration, we first analyzed the behavioral choice patterns of head-fixed mice trained on a dynamic foraging task. The data were

originally presented in (Hattori et al. 2019). In each trial, the mice were presented with the ready cue (light), followed 2-2.5 sec later by the answer cue (tone), after which they chose one of two options: lick left or lick right. There was no cue that instructs mice to choose one over the other, but the two lickports had different probabilities of delivering a water reward, schematized in Figure 3.1 A. These probabilities were stable for periods of time but changed every 60-80 trials, without any cue to the mouse. Mice trained in this task dynamically adjusted their choice pattern according to their choice-outcome history (example session, Fig. 3.1 B).

We quantified their use of the choice and outcome information from past trials with a logistic regression model. The model was fit using three types of history information: rewarded-choice history (the interaction between reward and choice, 1 for rewarded left choice, -1 for rewarded right choice, 0 otherwise), unrewarded-choice history (1 for unrewarded left choice, -1 for unrewarded right choice, 0 otherwise), and outcome-independent choice (1 for left choice, -1 for right choice, 0 otherwise), for recent trials. The rewarded-choice history influenced the behavior most strongly and we focused on this history for the rest of the study. The regression weights for individual history events (example session, Fig. 3.1 C) indicate that mice used the most recent history information more than distant history information, with significant weights for trials as far as 10 trials back.

The influence of past rewarded-choice experience decays smoothly on average (Fig. 3.1 D). Importantly, the shape of this decay exhibits a sharp initial drop on recent experience and a heavy-tail on more distant experience, which is better described by hyperbolic than exponential fit (exponential AIC: -32.28, hyperbolic AIC: -55.55; lower AIC indicates better fit). That is, distant history is weighted more than expected from a consistent decay across all

timesteps. This is notable because a hyperbolic-like decay is a feature of the behavior that standard RL models are unable to capture.

To confirm that behavior described by the RL model exhibits exponential decay, we generated an artificial choice pattern in an emulation of our task using a modified form of the RL model developed previously for this behavior (Hattori et al. 2019). The parameters of the RL model were taken from fitting the mouse behavior for each session. The sets of fitted parameters were then used with the generative model to produce simulated behavior, and the simulated choice patterns were fit with the same logistic regression model as above. By analyzing the history weights from the regression, we find that the simulated behavior is better fit by exponential than hyperbolic decay (Figure 3.1 E, exponential AIC: -56.33, hyperbolic AIC: -36.97), in contrast to the real behavior in Figure 3.1 D. Exponential behavior by the RL model is expected; the RL agent of the simulation uses a recursive style of integration that is time-invariant and therefore by definition exponential in nature. This result confirms that our analysis can accurately detect this feature.

The results so far, based on the history weights from regression fits, suggest that the mice are using a hyperbolic-like integration rather than exponential integration to make their decision on a trial-by-trial basis. We tested this more directly by comparing the fits of two models where decay functions were convolved directly with the rewarded-choice pattern, rather than fit to the regression weights post hoc (Fig. 3.1 F). We constructed this model with the explicit constraint that past weights decay monotonically with either an exponential or hyperbolic decay function. We then assessed whether the model with an exponential or hyperbolic constraint better fit the observed behavior. Model fit was evaluated by the session-by-session difference in cross-validated loglikelihood of hyperbolic and exponential models.

In this nomenclature, a likelihood difference larger than zero indicates that a hyperbolic constraint better fits the behavior, and less than zero that exponential fits better. We observe that real behavior is in fact better fit by the hyperbolic model (Fig. 3.1 G, median=0.34, p=1.6e-11, 2-tailed Wilcoxon sign rank). In contrast, the simulated behavior generated with the RL was better fit with exponential integration (Fig. 3.1 G, median=-0.24, p=2.3e-8, 2-tailed Wilcoxon sign rank), as expected.

These results establish that the mice are using a behavioral strategy that deviates from the standard RL model, integrating history information with hyperbolic-like decay function.

Figure 3.1: Mice rely on hyperbolic rather than exponential weighting of rewarded-choice history during history-dependent value-based decision-making.

(A) Schematic of behavior task. Mouse is presented with two lickspouts with different probabilities of reward on the left or right side. The mouse was cued with an amber LED to withhold licking during the ready period, then cued with a tone to choose a side in the answer period. The reward contingency inverted in a block structure of variable block lengths, and the pattern was repeated until the end of the session. The first block was randomly selected to be right- or left-high for any session.

(B) Example session, probability of left reward assignment (black line), 10-trial smoothed choice pattern (purple line), left and right licks (blue, red), that were rewarded or unrewarded.

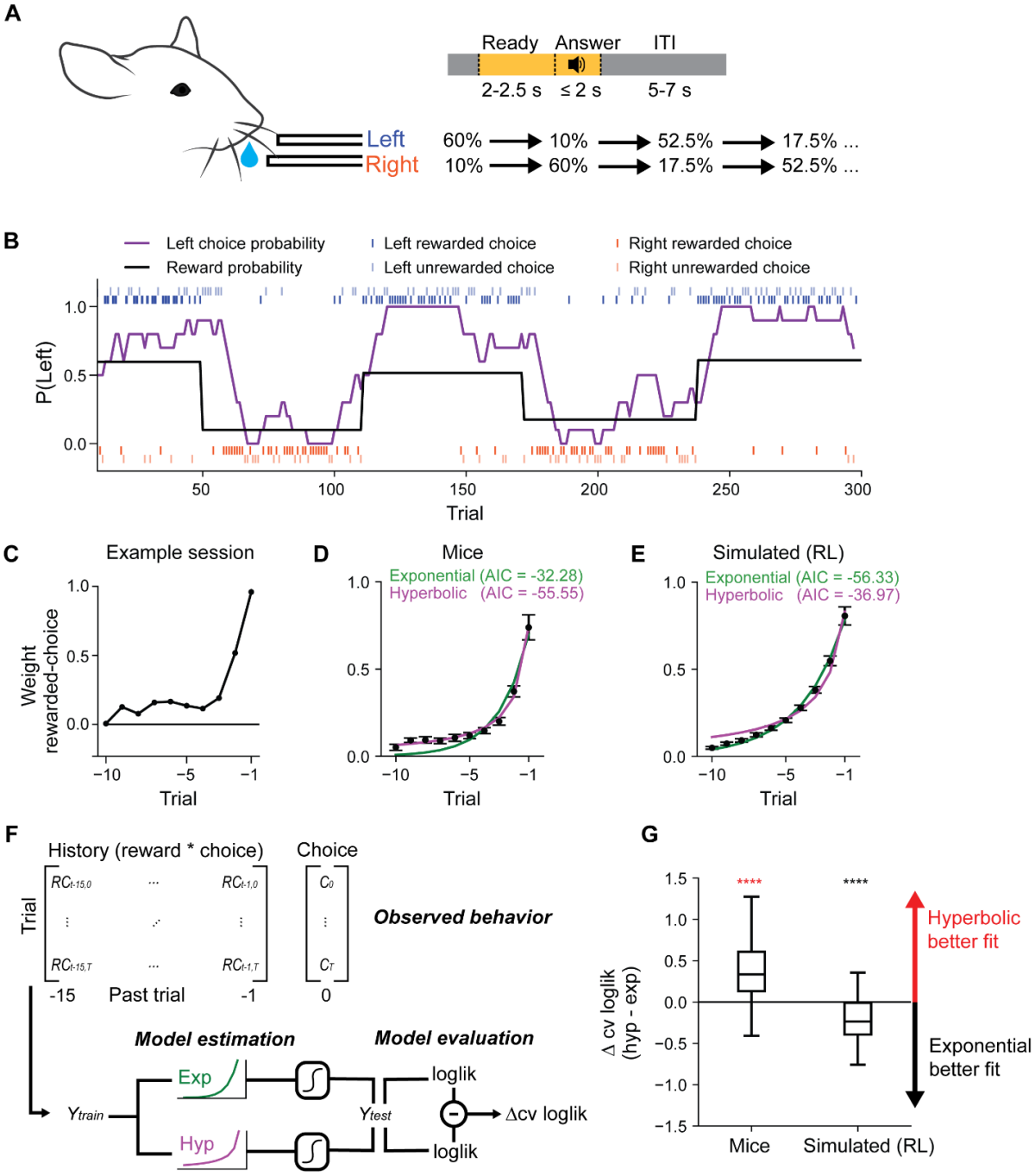
(C) Logistic regression weights on rewarded-choice history for the example session in B.

(D) Rewarded-choice weights from logistic regression in black, grand mean across 74 sessions and 14 animals (mean ± SEM). Exponential (green) and hyperbolic (magenta) curves fit to the mean weights; exponential AIC: -32.28, hyperbolic AIC: -55.55; lower AIC indicates better fit.

(E) As in D, but for 74 simulated sessions with unique input parameters (mean + SEM). Exponential AIC: -56.33, hyperbolic AIC: -36.97; lower AIC indicates better fit.

(F) Analysis workflow of the exponential and hyperbolic behavioral integration models.

(G) Comparison of model performance, using 10-fold cross-validated loglikelihood, compared between exponential and hyperbolic models across identical train- and test-sets. Red indicates median above zero, black median below 0. (Mice: p=3.24e-11, simulated: p=2.28e-8, 2-tailed Wilcoxon signed-rank, FDR corrected for multiple comparisons).


A
Ready
Answer
ITI
2-2.5 s
≤ 2 s
5-7 s
Left
Right
60% → 10% → 52.5% → 17.5% ...
10% → 60% → 17.5% → 52.5% ...
B
Left choice probability
Reward probability
Left rewarded choice
Left unrewarded choice
Right rewarded choice
Right unrewarded choice
P(Left)
Trial
C
Example session
Weight rewarded-choice
Trial
D
Mice
Exponential (AIC = -32.28)
Hyperbolic (AIC = -55.55)
Trial
E
Simulated (RL)
Exponential (AIC = -56.33)
Hyperbolic (AIC = -36.97)
Trial
F
History (reward * choice)
Choice
Trial
Past trial
Observed behavior
Model estimation
Model evaluation
Exp
Hyp
Ytrain
Ytest
loglik
Δcv loglik
G
Δ cv loglik (hyp - exp)
Hyperbolic better fit
Exponential better fit
Mice
Simulated (RL)

3.2.2 Cortical neurons encode rewarded-choice history with exponential-like integration

To explore the neural basis of hyperbolic history integration, we analyzed neural activity recorded from task-performing mice. These data, originally described in (Hattori et al. 2019), were acquired with in vivo two-photon calcium imaging in CaMKIIa-tTA::tetO-GCaMP6s double transgenic mice expressing GCaMP6s in cortical excitatory neurons (Fig. 3.2 A). Fluorescence traces from each neuron were deconvolved (Friedrich, Zhou, and Paninski 2017; Pachitariu, Stringer, and Harris 2018) to give an approximation of underlying spiking activity. We focused our analysis on 5 cortical areas; retrosplenial cortex (RSC), posterior parietal cortex (PPC), posterior premotor cortex (pM2), anterior lateral motor cortex (ALM), and primary somatosensory cortex (S1).

The activity of a subset of cortical neurons was modulated by rewarded-choice history. As seen in three example cells imaged in the same session in RSC, shown in Fig. 3.2 B, these cells exhibited different levels of activity depending on whether the left or right choice was rewarded in recent trials. The clearest separation in activity was when the most recent trial was rewarded on either the left side (darkest blue) or the right side (darkest red). Some of these cells showed stronger activity following left rewarded choice (e.g. cell 2), while others following right rewarded choice (e.g. cells 1 and 3). We focused the following analysis on the activity during the pre-choice, ready period (2 sec after the ready cue onset). We quantified the fraction of neurons modulated by rewarded-choice on at least the most recent trial (trial-1, Fig. 3.2 C), using linear regression (methods eq. 10). The fraction of significantly modulated neurons varied across sessions and across cortical areas, but was always well above chance, as calculated by shuffling the neural activity across trials.

To investigate how these history-modulated neurons perform history integration, we applied the analogous model as for the behavior. Specifically, to each cell we fit a pair of

models in which past choice history was constrained to display either an exponential or hyperbolic decay and quantified the model's prediction of cell activity with cross-validated loglikelihood. In contrast to the mouse behavior, we found that the cell activity was generally better fit by the exponential integration model (Fig. 3.2 D, see also Fig. 3.1 G). The cells were more exponential than hyperbolic across all cortical areas we investigated (RSC: median=-3.19, $p<<1e-10$; PPC: median=-2.65, $p<<1e-10$; pM2: median=-2.40, $p<<1e-10$; ALM: median=-2.26, $p<<1e-10$; S1: median=-1.78, $p<<1e-10$).

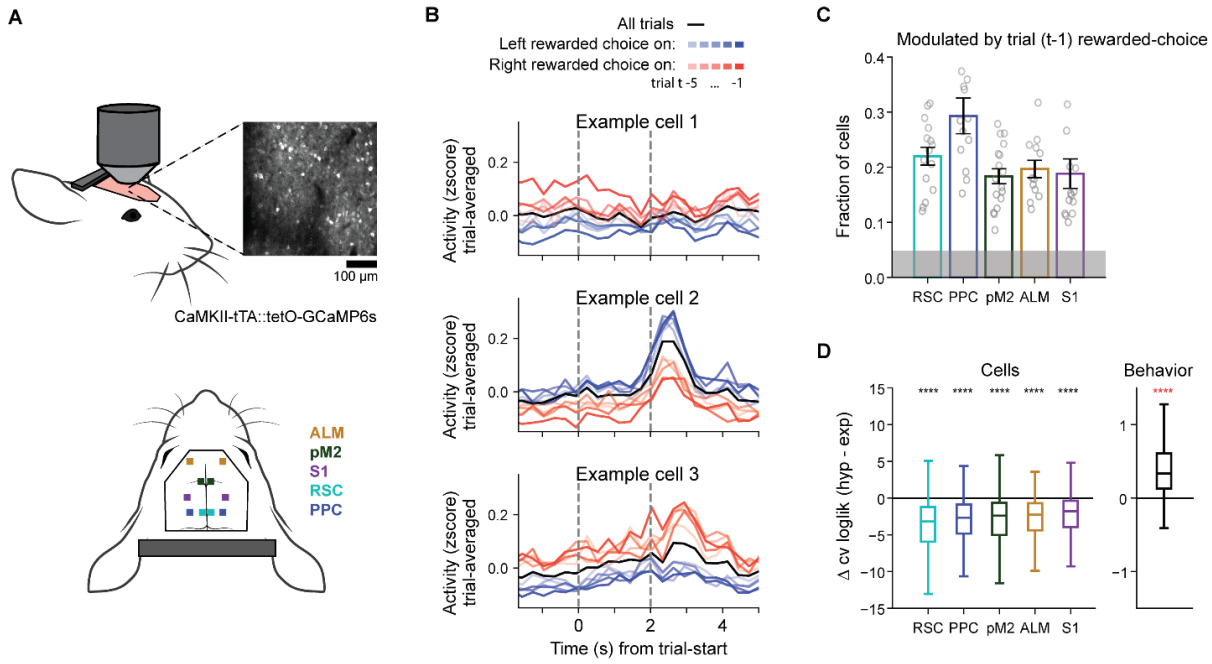


Figure 3.2: Cortical neurons encode history information with exponential decay.

(A) Schematic of two-photon imaging from five cortical areas, showing one example field of view from RSC. One cortical area in one hemisphere was imaged per session.

(B) Trial-averaged activity of three example RSC neurons. Black line is the average across all trials. Blue lines are the mean of the subset of trials where the past trial (-5:-1 trials, indicated by darkening shade) was left choice and rewarded. Red lines are the same, but for right choice and rewarded. Grey dashed lines indicate the first 2 seconds of the 'ready' period of the trial.

(C) Fraction of cells significantly modulated by rewarded-choice on the most recent trial (t-1). (mean ± SEM). Grey shading indicates fraction of significant cells in trial-shuffled data. (n sessions: RSC=15; PPC=16; pM2=17; ALM=12; S1=14).

(D) Comparison of model performance, using 10-fold cross-validated loglikelihood, between exponential and hyperbolic models across identical train- and test-sets. Left: loglikelihood from the regression model fit to the cell activity; right: loglikelihood of the generalized linear model fit to the behavior. Behavior replicated from 1G for comparison. Note that loglikelihood for cell model and behavior model cannot be directly compared. (2-tailed Wilcoxon signed-rank, ****, $p < 0.0001$, FDR corrected for multiple comparisons).

3.2.3 Cortical neurons encode temporal information with a wide variety of time-constants

How can the brain generate behavior with hyperbolic integration when cortical neurons demonstrate exponential decay of past information? We consider the possibility that a hyperbolic discounting function with a sharp initial decay and a heavy tail can be approximated by a combination of exponentials with short and long time-constants. Therefore, if cortical neurons perform exponential integration with the decay time-constants that are heterogeneous across neurons, their combination could lead to a hyperbolic-like function to guide behavior. This mechanism for the generation of hyperbolic behavior from exponential neurons requires that there be a sufficiently diverse pool of neural decay rates to provide a basis for hyperbolic behavior. To test this idea, we examined the exponential decay time-constants of history modulated cells. Indeed, we observed that even within one field-of-view for one cortical area, there are a wide variety of decay rates across cells (example cells from an RSC session, Fig. 3.3 A). Take, for example, cell 1, which shows the sharpest convergence between the cell activity traces (Fig. 3.2 B, top), corresponding to a short integration time (Fig. 3.3 A, top). In contrast, example cell 3 still showed clear separation of activity traces dependent on a rewarded-choice as many as five trials back (Fig. 3.2 B, bottom), leading to the more slowly-decaying exponential fit in Figure 3.3 A, bottom. To investigate the distribution of decay time-constants across neuronal populations, we focused our analysis on cells that were significantly modulated by rewarded-choice history in both first and second halves of the session. These stable, exponentially-modulated cells represent a smaller fraction than the cells which are modulated at least by the most previous rewarded choice for some part of the session, but these fractions were well above chance from the shuffled distribution in all 5 areas (Fig. 3.3 B). RSC and PPC exhibited the highest fraction of

modulated cells, while pM2, ALM, and S2 all showed less prevalent encoding of rewarded-choice history information (compared to RSC, PPC: p=0.80, pM2: p=2.16e-4; ALM: p=1.24-4; S1: p=3.61e-4).

We observed that these exponential cells show a wide variety in their decay time-constants (Fig. 3.3 C). Interestingly, the distributions of time-constants differed across areas. RSC was particularly enriched in neurons that encode history information with longer time-constants: there was a right-shift in the distribution of tau in RSC compared to the other cortical areas (median, RSC: 2.34; PPC: 1.74; pM2: 1.76; ALM: 1.31; S1: 1.68. Compared to RSC, PPC: p=3.2e-04; pM2: p<1e-5; ALM: p=1.6e-04, S1: p=8.0e-05; bootstrapped test of medians, p-values Bonferroni corrected).

To confirm that these distributions are not simply due to estimation noise, we partitioned each session into two blocks of trials and evaluated the exponential time-constants separately in each block for each neuron. As shown in the example RSC session in Fig. 3.3 D, the time-constants estimated in two halves of the session for the same neuron were highly consistent (Spearman's r=0.58, p=1.54e-8). Across cortical areas, we routinely found that cell-specific tau was consistent, with correlations substantially higher than trial-shuffled data (Fig. 3.3 E; real data: mean r=0.47, geometric-mean p=2.05e-5; shuffled data: mean r=-0.01, geometric-mean p=0.31). This consistency indicates that each neuron integrates history with a stable time-constant that is specific to the cell and consistent throughout a session, and that this time-constant can be reliably estimated.

Having established that multiple temporal scales are encoded simultaneously across neural populations in cortex, next we considered how the observed distributions of time-constants, which differed across areas, could relate to the behavioral strategy. Specifically, we

asked whether a weighted sum of these diverse exponential functions could approximate the hyperbolic-like integration observed in behavior. To answer this, we designed a linear regression model in which behavioral choice patterns of mice were fit by the weighted sum of multiple exponential integrators with different time-constants. The time-constants were randomly sampled from those observed in cortical neurons, and the weights associated with each of the time-constants fit to the behavior (schematized in Fig. 3.3 F). We quantified the fit of this model with weighted sum of exponentials, which we call 'quasi-hyperbolic', as the difference in the cross-validated loglikelihood from the performance of the model with a single exponential function with the best-fit time-constant (analysis outlined in Fig. 3.3 G). We varied the number of time-constants for the quasi-hyperbolic model, and for each number of time-constants the random sampling of time-constants was repeated 1000 times and the results were averaged for each session. As we increased the number of sampled time-constants, the performance of the quasi-hyperbolic model improved, surpassing that of the best-fit single exponential model and converging to the performance of the hyperbolic model (Fig. 3.3 H, grey line being the average improvement of the hyperbolic model from Fig. 3.1 G). The performance of the quasi-hyperbolic drawn from the RSC neurons improved most quickly with an increasing number of sampled time-constants, indicating that RSC temporal encoding best matches the temporal characteristics of the behavior. Put another way, a downstream read-out of information from RSC can reproduce the observed timescale of the behavior more parsimoniously than any other area. This suggests that RSC holds a unique position among these cortical areas as having a representation of temporal history information that best matches the temporal component of the behavior.

Figure 3.3: Cell time-constants are heterogeneous, and the time-constants in RSC cells best match the behavior.

(A) Mean activity difference (left-right) taken from the 'ready' period of Fig. 3.2 B. Black dots are the difference between left-rewarded activity and right-rewarded activity (mean ± SEM). Exponential filter (green line) estimated by the model for each cell. Cell 1: tau=0.81; cell 2: tau=1.45; cell 3: tau=4.35.

(B) Fraction of cells significantly modulated by rewarded-choice history with an exponential decay in both halves of the session (mean ± SEM). Open bars are the fraction of cells modulated by only the most recent trial (t-1) reproduced from Figure 3.2 C. Grey shading indicates fraction of cells significantly modulated in trial-shuffled data. Cell fractions compared to RSC by two-tailed paired t-test (PPC: p=0.57, pM2: p=9.21e-4; ALM: p=5.83e-4; S1: p=3.78e-4, with FDR corrected for multiple comparisons). (n sessions: same as Figure 3.2).

(C) Distribution of exponential time-constant τ across the significantly-modulated cells in five cortical areas. Top: histograms on log axis. Bottom: boxplots on linear axis. All sessions for given area are pooled. Medians are RSC: 2.62; PPC: 1.54; pM2: 1.62; ALM: 1.16; S1: 1.36. (Bootstrapped test of medians compared to RSC, ****, $p < 0.0001$; p-values FDR corrected for multiple comparisons).

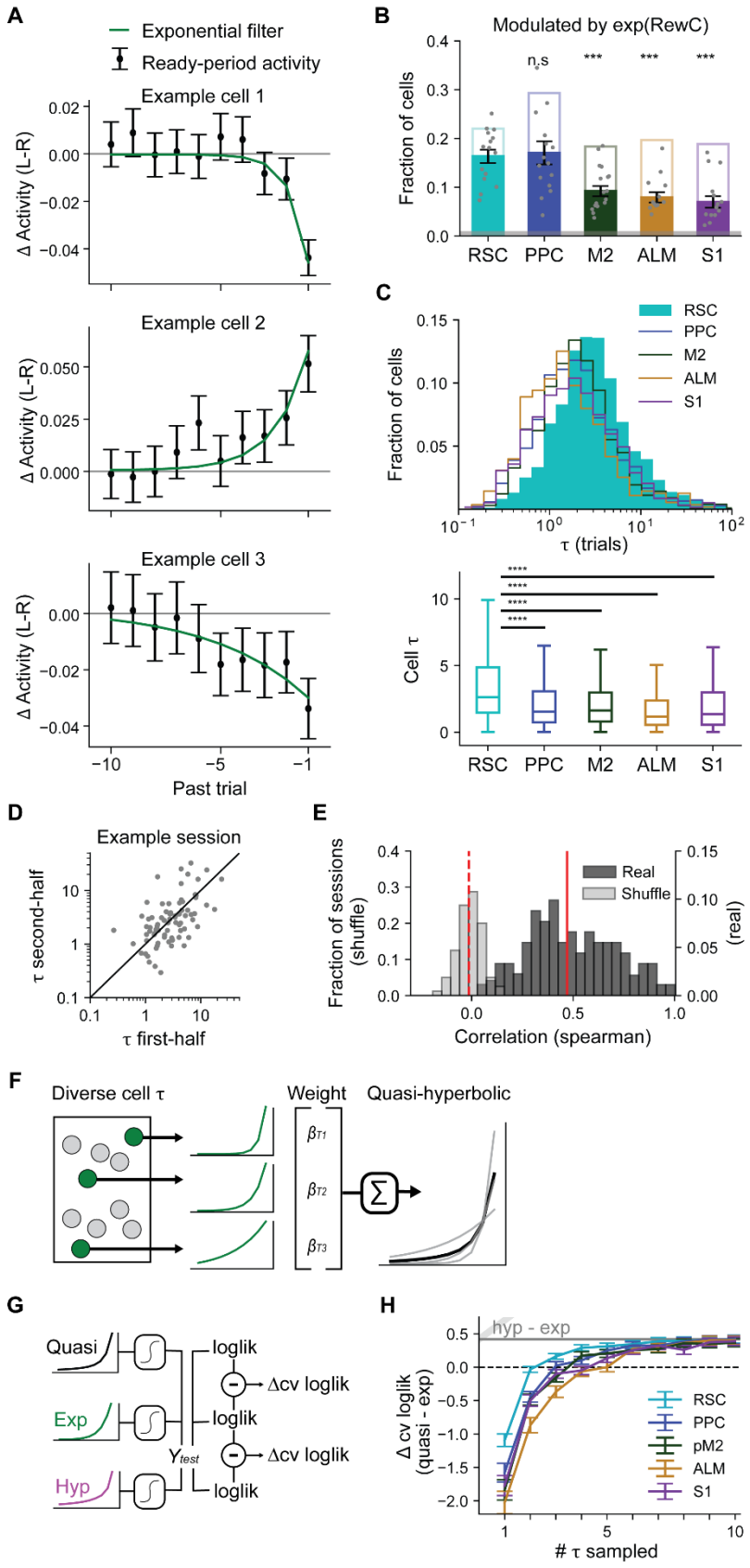
(D) τ estimated separately in two halves of the session. Spearman's r=0.58, p=1.54e-8.

(E) Distribution of the Spearman's correlation across all session-splits. All areas combined. Real data: mean r=0.47 (red line), geometric-mean p=2.05e-5; shuffled data: mean r=-0.01 (dashed red line), geometric-mean p=0.31.

(F) The quasi-hyperbolic model is the weighted sum of multiple exponential processes, which yields a heavy-tailed function approximating a hyperbolic.

(G) Model performance is compared as the difference in cross-validated loglikelihood between the quasi-hyperbolic model and the exponential model, compared to the improvement of the hyperbolic model over the exponential model.

(H) Performance of the quasi-hyperbolic behavioral model, compared to exponential behavioral model. Mean ± SEM across 1000 random draws. Grey line is mean ± SEM of hyperbolic model, reproduced from Figure 3.1 G.


A
Exponential filter
Ready-period activity
Example cell 1
Example cell 2
Example cell 3
Δ Activity (L-R)
Past trial
B
Modulated by exp(RewC)
Fraction of cells
n.s
***
RSC
PPC
M2
ALM
S1
C
Fraction of cells
τ (trials)
Cell τ
****
D
Example session
τ second-half
τ first-half
E
Fraction of sessions (shuffle)
Real
Shuffle
(real)
Correlation (spearman)
F
Diverse cell τ
Weight
Quasi-hyperbolic
G
Quasi
Exp
Hyp
loglik
Δcv loglik
Ytest
H
hyp - exp
Δ cv loglik (quasi - exp)
pM2
τ sampled

3.2.4 Inactivating RSC, but not PPC or pM2, reduces the mouse's use of rewarded-choice history in hyperbolic-like integration

In the above results, we laid out evidence that history integration occurs at multiple timescales simultaneously across different neurons in cortex, and RSC is enriched in the timescales that best match the behavior. From this, we hypothesize that RSC is uniquely required for the history integration to guide the behavior.

To test this, we selectively and reversibly inactivated cortical areas via optogenetic activation of Parvalbumin-positive (PV) inhibitory neurons in PV-Cre::LSL-ChR2 double transgenic animals that expressed Channelrhodopsin in PV neurons. We focused on RSC, PPC, and pM2, the three areas with strongest history encoding. We used a projector system (Hattori et al. 2019; Dhawale et al. 2010; Haddad et al. 2013) to apply blue light over each cortical area. This flexible light-delivery system, combined with a large cranial window preparation (Hattori and Komiyama 2022c), allowed us to investigate the role of multiple cortical areas separately within the same animal (schematic in Fig. 3.4 A). Inactivation was performed only for one area per session. In each session, inactivation occurred on a subset of trials (15% randomly selected, with the constraint to not be within 3 trials of each other), starting from the beginning of the ready cue until the choice was made. On all other trials light was directed over the headbar, away from the brain, in the same task period to control for light distraction. We also performed separate control sessions in which the light was applied over the headbar in all trials and designated ~15% of these trials as pseudo-inactivation trials. The total area of light coverage and intensity per cortical area was consistent for each cortical area, and for the control condition. 3 of the x RSC inactivation animals have been previously described (Hattori et al. 2019). Inactivation on one trial could

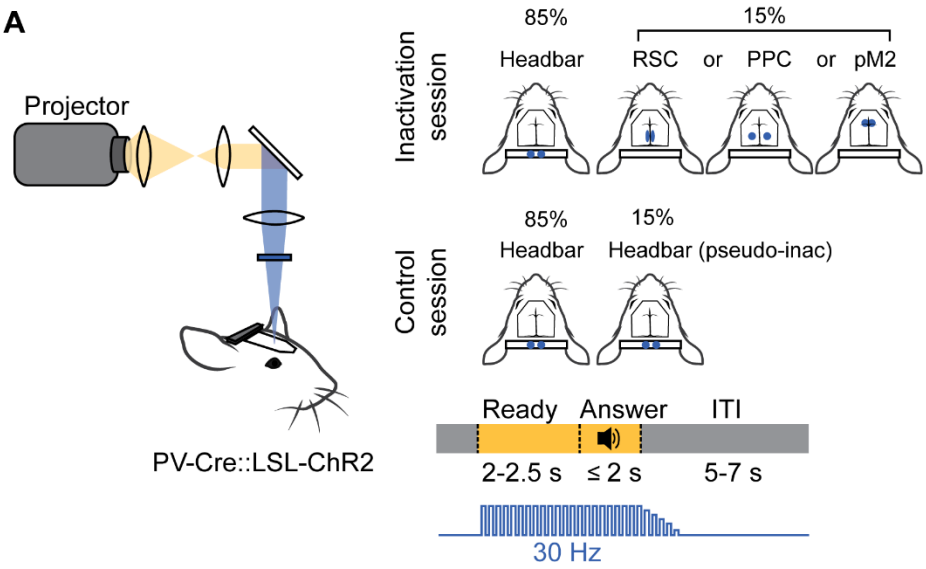
have carry-over effects on subsequent trials. As such, we compared the inactivation trials with pseudo-inactivation trials in the control sessions of the same animal.

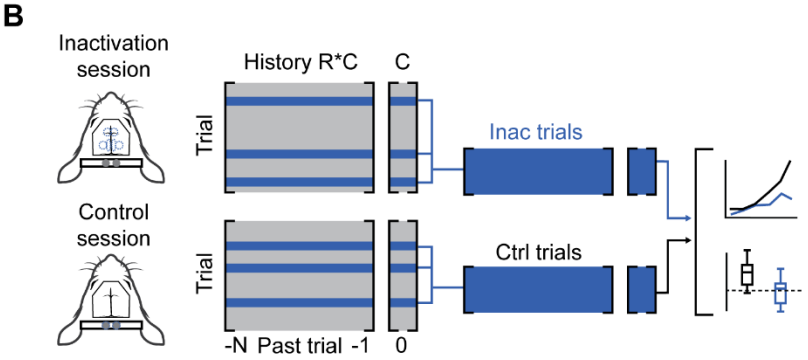
To quantify the effects of inactivation, we fit a modified version of the logistic regression analysis (methods eq. 3-4) in which the inactivation trials had a separate set of weights from the pseudo-inactivation trials from control sessions (analysis outline in Fig 3.4 B.). We found that inactivating RSC during the pre-choice, ready period reduced the dependence on rewarded-choice history. This effect was not seen with inactivation of PPC or pM2(Fig. 3.4 C, effect of inactivation condition: RSC: p=3.15e-4; PPC: p=0.90; pM2: p=0.37, two-way repeated-measures ANOVA). Thus, of these three areas with strong representation of history information, we found that only RSC is uniquely necessary for the behavioral use of rewarded-choice history.

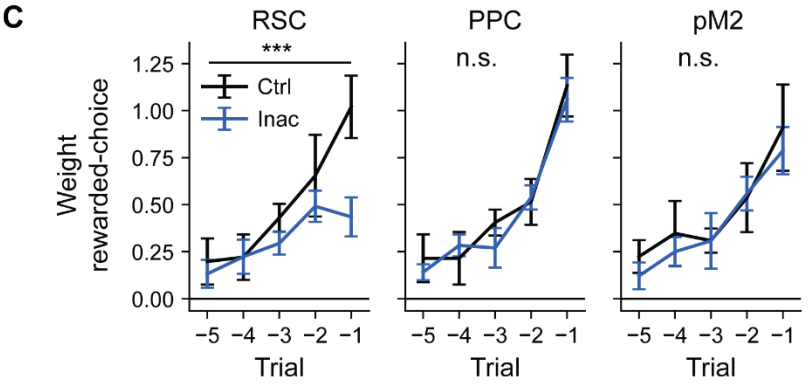
Lastly, we investigated whether inactivation affected the hyperbolic nature of behavioral history integration. The behavioral decay models, hyperbolic and exponential, were estimated separately for the inactivation trials or pseudo-inactivation trials, and model performance calculated as the difference in cross-validated loglikelihood between hyperbolic and exponential models. We found that RSC inactivation caused the behavioral strategy to become less hyperbolic than in control trials. This effect was not seen with inactivation of PPC or pM2 (Fig. 3.4 D, RSC: p=6.06e-3; PPC: p=0.52; pM2: p=0.91). These results indicate that RSC is necessary for implementing the hyperbolic-like integration we observe in the choice patterns.

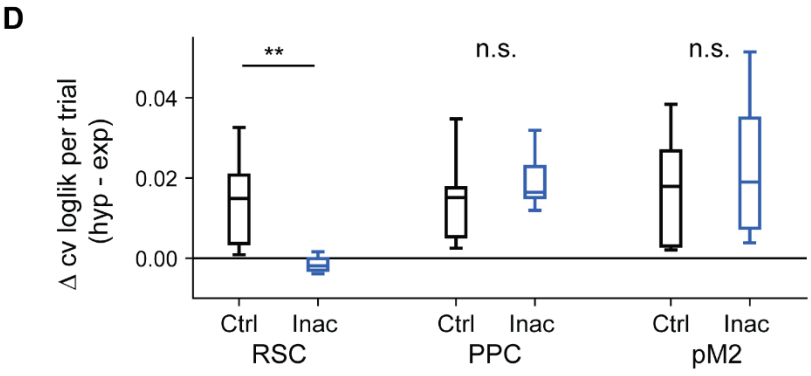
Figure 3.4: Inactivation of RSC reduces reliance on rewarded-choice history and impairs hyperbolic weighting of past trials.

(A) Schematic of the inactivation. Patterns of light were delivered with a projector-based system onto the cortical surface of mice performing the task. Right: the position of stimulus for RSC, PPC, or pM2 during 15% of trials, one area per session, and to the headbar in the other 85% of trials. Control sessions included light over the headbar in all trials. Illumination at 30 Hz occurring during the ready and answer periods of trials.
(B) Analysis workflow for inactivation behavior data. The inactivation trials for inactivation sessions and 15% of trials in control sessions are considered in the analysis.
(C) Logistic regression weights in control (Ctrl, black line) and inactivation (Inac, blue line) trials. (mean ± SEM) (RSC: n=10 mice, 29 control sessions, 26 opto sessions; PPC: n=10 mice, 31 control sessions, 26 inactivation sessions; pM2: n=9 mice, 28 inactivation sessions, 22 inactivation sessions). (Two-way repeated-measures ANOVA, effect of inactivation condition. RSC: p=3.15e-4; PPC: p=0.90; pM2: p=0.37).
(D) Comparison of model performance, using 10-fold cross-validated loglikelihood, compared between exponential and hyperbolic models across identical train- and test-sets. (2-tailed Wilcoxon signed-rank. RSC: p=6.06e-3; PPC: p=0.52; pM2: p=0.91).


A
Projector
PV-Cre::LSL-ChR2
Inactivation session
85%
Headbar
15%
RSC or PPC or pM2
Control session
85%
Headbar
15%
Headbar (pseudo-inac)
Ready
Answer
ITI
2-2.5 s
≤ 2 s
5-7 s
30 Hz


B
Inactivation session
Control session
History R*C
C
Trial
Inac trials
Ctrl trials
-N Past trial -1
0


C
RSC
PPC
pM2
***
n.s.
n.s.
Ctrl
Inac
Weight rewarded-choice
1.25
1.00
0.75
0.50
0.25
0.00
−5 −4 −3 −2 −1
Trial


D
**
n.s.
n.s.
Δ cv loglik per trial (hyp - exp)
0.04
0.02
0.00
Ctrl
Inac
RSC
PPC
pM2

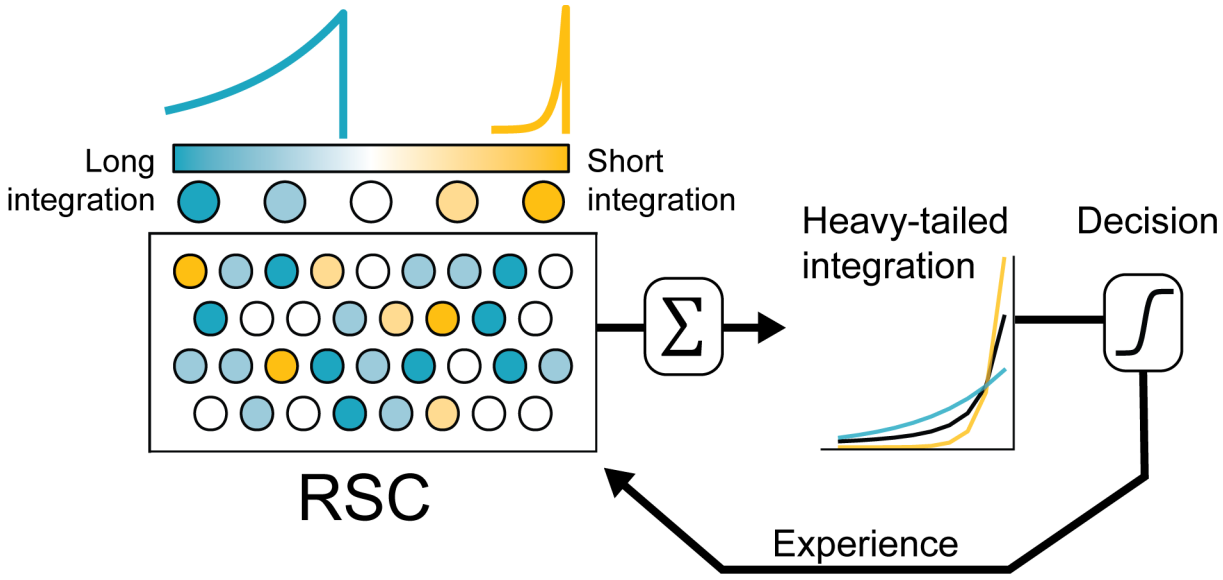


Figure 3.5: RSC encodes a reservoir of temporal information used for decision-making. RSC neurons encode rewarded-choice history experience with a diversity of time-constants, including cells with short integration and long integration. The combination of many exponential integrators yields the heavy-tailed integration observed in the behavior.

3.3 Discussion

Here we showed that the behavior of mice engaged in value-based decision-making is driven by trial history integrated according to a hyperbolic-like decay, similar to what has been shown in analogous behavioral tasks in multiple species (Serences 2008; Sugrue 2004; Corrado et al. 2005; Lau and Glimcher 2005; Aparicio and Baum 2009; Iigaya et al. 2019). In an apparent contradiction, we found that cortical neurons encode history in a manner more consistent with an exponential decay than hyperbolic. However, cortical neurons do not represent history homogeneously. Rather, history is encoded simultaneously across many neurons with heterogenous time-constants of integration. This activity pattern is consistent with a series of exponential processes acting in parallel over widely distributed temporal horizons, which can sum together to yield the heavy-tailed, hyperbolic-like integration observed in the behavioral strategy. RSC encodes this information over a longer temporal horizon than the other cortical areas, and inactivation of RSC uniquely attenuates the use of history information and impairs the hyperbolic-like integration. From these results, we posit that history information is integrated in a distributed and diverse manner, with experience across different timescales encoded differently across neurons. We propose a model where history information is encoded in individual neurons by simple exponential integration with heterogeneous time-constants. Behavior arises from the combination of multiple exponentially-integrative processes, in particular those with longer time-constants, which yields a decision strategy that has the heavy-tailed feature of a hyperbolic (Fig. 3.5).

This model provides a potential mechanism by which the brain uses a conceptually simple mechanism of exponential value updates to achieve a hyperbolic-like behavior. Such a heavy-tailed integration has also been described as the phenomenon of ‘undermatching’, and has previously been considered a suboptimal form of decision-making in laboratory behaviors

(Corrado et al. 2005; Lau and Glimcher 2005; Iigaya et al. 2019). However, a natural environment has dynamics at multiple timescales. For example, the decision of when and where to go foraging for food may depend on factors such as the weather, hunger state, and seasonal changes in available items, just to name a few. These factors vary across orders of magnitude in the speed of changes, rendering an exponential integration with a constant decay rate per unit time to be suboptimal. Hyperbolic-like decay is conserved across species, and may be an evolutionary response to an environment that has multiple timescales of changes. Although here we focused on retrospective processes in which history information decays over time, prospective discounting of potential future rewards is also known to follow a hyperbolic function. For example, in delay discounting experiments in which the animal is presented with choices that return rewards at different temporal delays and magnitudes, animals show a stereotypic reversal in time-preference at long-lags indicative of hyperbolic discounting (Frederick and Loewenstein 2002; Haith, Reppert, and Shadmehr 2012). Thus, hyperbolic functions seem common in weighting influences of events over time. However, truly hyperbolic computation is difficult to achieve with a recursive operation modeled in standard RL, and other computations that approximate hyperbolic discounting have been proposed (Alexander and Brown 2010; Fedus et al. 2019; Kurth-Nelson and Redish 2009; Wilson, Nassar, and Gold 2013). In contrast, exponential computation can be achieved with a simple recursive operation. In an extension of this logic, our model proposes that multiple exponential computations with heterogeneous time-constants performed in parallel can generate hyperbolic-like behavior.

Previous studies have also described heterogeneity across individual neurons in their encoding of behavior-related temporal information within and across brain areas (Brody 2003;

Goldman 2017; Harvey, Coen, and Tank 2012; Bernacchia et al. 2011; Spitmaan et al. 2020; Scott et al. 2017; Runyan et al. 2017). We extend these observations and uncover that RSC is especially enriched in long time-constants, and that the timescales in RSC best match the animal's behavioral strategy. RSC inactivation leads to an impairment in the animal's ability to use history information to make its decision. Furthermore, our recent studies uncovered that RSC uniquely maintains history information as persistent population activity (Hattori et al. 2019; Hattori and Komiyama 2022a). These identify RSC as a critical cortical area that encodes and maintains behaviorally-relevant temporal information.

How do individual neurons perform diverse temporal integration? Even within the same area, cells are often heterogeneous in their intrinsic temporal characteristics (Murray et al. 2014; Cavanagh et al. 2016; Chaudhuri et al. 2015). However, such intrinsic time-constants of cells are short and not necessarily associated with their encoding of longer, reward-history relevant timescales (Spitmaan et al. 2020). Accordingly, the increasingly long temporal horizons represented in higher cortical areas likely also involve network interactions and reverberation (Tiganj, Hasselmo, and Howard 2015; Chaudhuri et al. 2015; Hunt and Hayden 2017; Spitmaan et al. 2020). Notably, RSC has been found to have a neural population dominated by highly excitable cells, yielding a network capable of producing sustained and specific responses to outside input (Brennan et al. 2020). Indeed, during value-based decision making, the activity of RSC neurons is particularly persistent, indicating the capacity for sustained encoding of behavior-relevant information (Hattori et al. 2019; Hattori and Komiyama 2022a).

When the environment shifts dynamically between periods of stability and volatility, animals need to adapt their behavioral integration timescales (Dayan, Kakade, and Montague

2000; Daw et al. 2006; Daw, Niv, and Dayan 2005; Daw et al. 2011; Kennerley et al. 2006; Behrens et al. 2007; Meder et al. 2017; Massi, Donahue, and Lee 2018; Bernacchia et al. 2011; Iigaya et al. 2019). In theory, RSC neurons could serve as a stable reservoir of heterogeneous history information, and behavioral adaptation could be achieved by flexibly adjusting the readout weights for different neurons by the downstream circuit. Such a shift might be mediated by neuromodulators whose activity is correlated with behavioral adaptation, such as dopamine (Doya 2002; Ito and Doya 2009; Kim et al. 2020), noradrenaline (Doya 2002; Yu and Dayan 2005), or serotonin (Doya 2002; K. W. Miyazaki et al. 2014; Cohen, Amoroso, and Uchida 2015; Matias et al. 2017; K. Miyazaki et al. 2020; Grossman, Bari, and Cohen 2022). Alternatively, individual RSC neurons may alter their time-constants of integration according to changes in environmental demands. The dynamics of RSC encoding and downstream readout during behavioral adaptations would be an interesting topic of future studies.

Our results resonate with distributional reinforcement learning (Dabney et al. 2020; Lowet et al. 2020), which models decision as being made as the combination of parallel estimates of value that vary in their degrees of optimism and learning rates. The advantage of such a system is to simultaneously provide information to the animal about the range of expected outcomes, forming a distribution of prospective reward expectation. Supporting this notion, midbrain dopaminergic neurons represent reward prediction error with diverse reversal points between positive and negative prediction errors (Dabney et al. 2020). As temporal integration may be sped or slowed with larger or smaller learning rates on reward-prediction errors, our model of hyperbolic behavior achieved from diverse speeds of neuronal

integration could be consistent with the theoretical model of distributional reinforcement learning, as either a consequence or cause of the value distribution.

Hyperbolic temporal integration confers a specific behavioral advantage, which is to balance sensitivity to both short-term changes and long-term trends. The distribution of temporal information observed in RSC is capable of producing the hyperbolic-like behavior, and our results suggest a specific cortical substrate and mechanism by which hyperbolic integration might arise.

3.4 Methods

3.4.1 Experimental Model and Subject details

All procedures were in accordance with protocols approved by the University of California San Diego Institutional Animal Care and Use Committee and the guidelines of the National Institutes of Health. The behavior and neural activity data from two-photon imaging were first reported in (Hattori et al. 2019), as was the RSC inactivation data for three of the twelve optogenetic inactivation animals. Both male and female mice were included in the imaging and inactivation datasets because we did not observe sex-related differences in their behavior or neural activity. Mice were originally purchased from the Jackson Laboratory. (CaMKIIa-tTA: B6;CBA-Tg(Camk2a-tTA)1Mmay/J [JAX 003010]; tetO-GCaMP6s: B6;DBA-Tg(tetO-GCaMP6s)2Niell/J [JAX 024742]; PV-Cre: B6;129P2-Pvalbtm1(cre)Arbr/J [JAX 008069]; Ai32: B6.Cg-Gt(ROSA) 26Sortm32(CAG-COP4* H134R/EYFP)Hze/J [JAX 024109]). All surgery, behavior training, and experiments were conducted in adult mice (6 weeks or older), on a reversed light cycle (12h-12h). Mice were water-restricted to ~1 ml/day while undergoing behavior training and experiments.

3.4.2 Surgery for imaging and optogenetics

Animals were prepared for imaging and optogenetic experiments with a large cranial window placed over dorsal cortex, as previously reported in (Hattori et al. 2019; Hattori and Komiyama 2022c). In brief, mice were anesthetized with 1-2% isoflurane during surgery, the dorsal surface of the skull was exposed and cleared of soft tissue with a razor blade, and marked with the coordinates of interest. After soaking the skull in saline, the bone became transparent enough to visualize the vasculature patterns on the surface of the brain. We took a photo with both the marked coordinates and vasculature visible, and used this as a reference to later identify the cortical area for two-photon imaging and inactivation. A large, hexagonal craniotomy was opened to expose all cortical areas of interest, and a glass window placed over the surface of the brain. The window was secured to the skull first with a thin application of 3M Vetbond (WPI), then with cyanoacrylate glue and dental acrylic cement (Lang Dental). Last, a custom-machined headbar was attached to the skull, posterior to the window using cyanoacrylate glue and dental cement. Mice were injected subcutaneously with dexamethasone (2 mg/kg) prior to surgery, and Buprenorphine (0.1 mg/kg) and Baytril (10 mg/kg) after surgery.

The cortical areas of interest for this study were anterior lateral motor (ALM, 1.7 mm lateral and 2.25 mm anterior to bregma), posterior premotor (pM2, 0.4 mm lateral and 0.5 mm anterior to bregma), posterior parietal (PPC, 1.7 mm lateral and 2 mm posterior to bregma), retrosplenial (RSC, 0.4 mm lateral and 2 mm posterior to bregma), and primary somatosensory (S1, 1.8 mm lateral and 0.75 mm posterior to bregma) cortex.

3.4.3 Behavior task

The dynamic foraging task and training paradigm were consistent with previous reports from the lab (Hattori et al. 2019; Hattori and Komiyama 2022a). In summary, mice were pretrained through a series of behaviors to introduce the task structure and to train licking to both left and right water delivery ports. In the foraging task, head-restrained mice were presented with two lickports monitored with IR beam detectors. Mice were required to withhold licking during a light-cued ready-period (2-2.5 sec) at the start of each trial, after which the mouse was cued with an auditory tone (10 kHz) to report a choice during the answer period (2 sec), after which they received a feedback tone (left: 5 kHz, right: 15 kHz), and probabilistic water reward. The water volume of reward was constant, at ~2.5 µl per reward. Following reward delivery, a variable-length inter-trial-interval followed (5-7 s), before the ready-period marked the beginning of the next trial. Trials in which the mice licked during the cued ready-period (‘alarm trials’) or trials in which the mouse did not make a choice during the answer period (‘miss trials’) were not rewarded, and excluded from analysis.

Reward was assigned to each lickport on every trial according to the reward probability for that lickport in that block. Once a reward was baited to a lickport it remained available there until chosen. The reward assignment probabilities for the two lickports were either [60%, 10%] or [52.5%, 17.5%]. This probability inverted randomly every 60-80 trials with a deterministic order of [60%, 10%], [10%, 60%], [52.5%, 17.5%], [17.5%, 52.5%], [60%, 10%], … with the first block being left-high or right-high at random.

3.4.5 Behavior inclusion criteria

For both the imaging and inactivation experiments, mice were trained over a period of weeks for 1-2 hours per session per day, to reach 'expert' level performance. Each session was evaluated for performance, and only consistent expert-level performance was included in analysis. Expert-level performance was assessed in part by the RL index (Hattori et al. 2019), which was a quantification of how closely the full RL model [eq. 5-7] captured the behavior. This was defined as the difference in model fit as follows:

$$RL\ index = \sqrt[n]{Likelihood\ of\ the\ RL\ model} - \sqrt[n]{Likelihood\ explained\ by\ bias\ only\ model} \quad [\mathbf{eq.\ 1}]$$

where the bias only model uses the bias term β_0 of the full RL model [eq. 7], and n is the number of choice trials in a session. Inclusion criteria consisted of an RL index of at least 0.08 for the session, and experience performing the task for at least 15 sessions. Expert mice usually performed > 600 trials in a session.

3.4.6 Two-photon calcium imaging and processing

As previously reported in (Hattori 2019), imaging experiments were conducted with a two-photon microscope (B-SCOPE, Thorlabs; 16x objective, 0.8 NA, Nikon) with excitation at 925 nm (Ti-Sapphire laser, Newport), continuously imaged at 29 Hz. Neurons were recorded from layer 2/3 in a single cortical area and hemisphere per session. We collected only one population from each hemisphere for each cortical area of a single mouse. The images were processed with a custom-built pipeline (Hattori and Komiyama 2022a) to correct motion artifacts (Mitani and Komiyama 2018) and image distortions (Hattori and Komiyama 2022b). We then used Suite2p (Pachitariu et al. 2016) to select cells and extract the GCaMP signal, identifying cells first with a user-trained classifier followed by manual inspection. This calcium signal was then deconvolved to obtain estimated spiking activity using a non-

negative deconvolution algorithm (Friedrich, Zhou, and Paninski 2017; Pachitariu, Stringer, and Harris 2018). This estimated activity for each neuron was z-score normalized across the time series for the entire session prior to all further analysis. The mean z score activity from the first 2 s of the ready-period, when the mouse is withholding licking, was used for the cell activity analysis.

3.4.7 Optogenetic inactivation

Cortical inactivation experiments were performed in PV-Cre::LSL-ChR2 double transgenic mice via activation of channelrhodopsin in the Parvalbumin-positive inhibitory neurons. Methods are consistent with (Hattori et al. 2019), and this paper includes RSC inactivation from three of the animals from the previous publication. Blue light was directed over the cortical surface through a large cranial window (described above for imaging) with a projector-based light delivery system. Elliptical illumination patterns were produced with Psychtoolbox in MATLAB, and projected (DLP projector, Optoma X600 XGA) through a single-lens reflex (SLR) lens (Nikon, 50 mm, f/1.4D, AF) coupled with 2 achromatic doublets (Thorlabs, AC508-150-A-ML, f = 150 mm; Thorlabs, AC508-075-A-ML, f = 75 mm) to focus illumination patterns over the brain and headbar. A dichroic mirror (Thorlabs, DMLP490L) and a blue filter (Thorlabs, FESH0450) were placed in the light path to pass only blue light (400-450 nm).

Cortical inactivation occurred on 15% of trials, constrained to not be within three trials of the previous inactivation. Light turned on at the beginning of the ready period, and turned off with the mouse's choice or the end of the answer period, whichever came first. During inactivation trials, light was directed over the cortical area of interest (one area per session), or over the headbar in control sessions. During control trials, light was directed over the headbar.

Light was pulsed at 30 Hz, at an intensity between 2.5-6 mW/mm2, with a linear ramp down of intensity at offset over 100 ms.

Three inactivation patterns were used: one for RSC, a 2.0 mm x 0.5 mm ellipse, centered at 0.3 mm lateral and 2.0 mm posterior to bregma; one for PPC, a 1.0 mm circle, centered at 1.5 mm lateral and 2.0 mm posterior to bregma; one for pM2, a 1.0 mm circle, centered 0.3 mm lateral and 0.5 mm anterior to bregma. Each pattern was bilaterally symmetric. The control light pattern was directed over the headbar as two 1.0 mm circles centered 1.0 mm apart. The stimulation pattern for each of the cortical and control conditions was light area- and intensity-matched.

3.4.8 Logistic regression behavioral model

To quantify the strategy of the mouse, we used a logistic regression model to predict the choices the mouse makes based on the recent experience the mouse has received. The choice on a given trial *t* is predicted by the weighted sum of the rewarded-choice (interaction of reward and choice, $RewC$), unrewarded-choice (interaction of no reward and choice, $UnrC$), and reward-independent choice (C) in the past 10 trials, along with a constant bias term. The model is:

$$logit\big(P_L(t)\big) = \sum_{i=1}^{10} \beta_{RewC(t-i)} * RewC(t-i) + \sum_{i=1}^{10} \beta_{UnrC(t-i)} * UnrC(t-i) + \sum_{i=1}^{10} \beta_{C(t-i)} * C(t-i) + \beta_0 \quad [\mathbf{eq.2}]$$

where $P_L(t)$ is the probability of choosing left on trial t, $RewC(t-i)$ is the rewarded choice on past trial $t-i$ (1 if rewarded on the left, -1 if rewarded on the right, 0 otherwise), $UnrC(t-i)$ is the unrewarded choice on past trial $t-i$ (1 if unrewarded on the left, -1 if unrewarded on the right, 0 otherwise), $C(t-i)$ is the reward-independent choice history on

trial $t-i$ (1 if left choice, -1 if right choice, 0 otherwise). $\beta_{RewC(t-i)}$, $\beta_{UnrC(t-i)}$, and $\beta_{C(t-i)}$ are the regression weights for each of the corresponding predictors, and β_0 is the constant bias term. Model fitting was performed in python, with the package Scikit-learn and the function *LogisticRegression*, solved by gradient descent with the BFGS algorithm.

3.4.9 Logistic regression for optogenetic analysis

Given that inactivation trials occurred on only 15% of trials in a session, we had a small number of trials for each session relative to the number of regression parameters in the above model. To increase model stability for these experiments, we reduced the number of history trials considered from 10 trials to 5 trials. Furthermore, instead of performing a regression for each session, we concatenated all inactivation and control sessions from each animal and used a version of mixed-effects model to account for variability across sessions. Only the inactivation trials and their corresponding history predictors were used, and coded as 'inac' for inactivation sessions, and 'ctrl' from the headbar control sessions. We further subsampled trials to have a matched number of trials in each condition, inac and ctrl, and iterated this subsample such that all trials were included at least once. The reported weights are the mean weights across all iterations.

The resulting model had the following fixed effects:

$$logit\big(P_L(t)\big) = \left(\sum_{i=1}^{5} \beta^{Ctrl}_{RewC(t-i)} * RewC(t-i) + \sum_{i=1}^{5} \beta^{Ctrl}_{UnrC(t-i)} * UnrC(t-i) + \sum_{i=1}^{5} \beta^{Ctrl}_{C(t-i)} * C(t-i) + \beta^{Ctrl}_0\right) * Ctrl(t)$$

$$+\left(\sum_{i=1}^{5} \beta^{Inac}_{RewC(t-i)} * RewC(t-i) + \sum_{i=1}^{5} \beta^{Inac}_{UnrC(t-i)} * UnrC(t-i) + \sum_{i=1}^{5} \beta^{Inac}_{C(t-i)} * C(t-i) + \beta^{Inac}_0\right)$$

$$* Inac(t) \qquad [\mathbf{eq.3}]$$

where the control, $Ctrl$, and inactivation, $Inac$, conditions had separate $\beta_{RewC(t-i)}$, $\beta_{UnrC(t-i)}$, and $\beta_{C(t-i)}$, and β_0 regression weights associated with them.

To estimate the random effects for individual sessions, the trials were assessed with random slope and intercept following the form:

$$y \sim f(t, Inac) + (0 + Inac|session) + (1|session) \quad [\mathbf{eq.4}]$$

Where $f(t, Inac)$ here refers to [eq. 3], $Inac$ is 1 on inactivation trials and 0 on control trials, with a random slope for the inactivation effect on each $session$, and a random intercept for each $session$.

3.4.10 Reinforcement learning model and simulated behavior

The reinforcement learning model used to generate simulated behavior is one taken from (Hattori et al. 2019). This model was modified from the Rescorla-Wagner model to describe mouse behavior in our task, with separately updated action values for the chosen option, Q_{ch}, and unchosen option, Q_{unch}:

$$Q_{ch}(t+1) = \begin{cases} Q_{ch}(t) + \alpha_{rew} * \big(R(t) - Q_{ch}(t)\big) \; if \; rewarded \; (R(t) = 1) \\ Q_{ch}(t) + \alpha_{unr} * \big(R(t) - Q_{ch}(t)\big) \; if \; unrewarded \; (R(t) = 0) \end{cases} \quad [\mathbf{eq.5}]$$

$$Q_{unch}(t+1) = (1 - \delta) * Q_{unch}(t) \quad [\mathbf{eq.6}]$$

where α_{rew} and α_{unr} are the independent learning rates for rewarded and unrewarded trials, respectively, and δ is the forgetting rate for the unchosen option. Reward on trial t is $R(t)$ (1 for rewarded, 0 for unrewarded), and the difference between $R(t)$ and Q_{ch} corresponds to reward prediction error (RPE). The learning and forgetting rates were constrained to be between 0 and 1. Given the action value for left and right options, which updated independently, the probability of choosing the left side is:

$$P_L(t) = \frac{1}{1 + e^{-\beta_{\Delta Q}(\beta_0 + Q_L(t) - Q_R(t))}} \quad [\mathbf{eq.7}]$$

where Q_L is value for the left side, Q_R is value for the right side, β_0 is the constant bias term, and $\beta_{\Delta Q}$ is the sensitivity to the value difference ΔQ. This model was fit to the choice patterns of 74 sessions of expert mouse behavior in python using SciPy *minimize* function, with search algorithm L-BFGS-B, to perform maximum likelihood estimation.

In an emulation of the task environment, with the same reward contingencies and block structure as the real task, the RL model algorithm was used to generate choices based on the trial-by-trial updating value from [eq. 5-6] and the soft max function [eq. 7]. This generative model took as inputs the fit parameters from each session of expert mouse behavior. The simulated RL agent ran 10,000 trials for each of the 74 parameter sets, producing sequences of choices and outcomes. These choice and outcome patterns were then fit with the logistic regression or behavioral models in the identical analysis process as real behavior.

3.4.11 Exponential and Hyperbolic behavioral models

To evaluate how well the mouse behavior is described by exponential or hyperbolic integration, we quantified the behavior using two explicitly-defined decay models that assumed either exponential or hyperbolic decay. Past trials were temporally discounted with an exponential or hyperbolic decay, with time-constants fit for each session.

The exponential model was defined as:

$$logit\big(P_L(t)\big) = \beta_{RewC} * \sum_{i=1}^{N} RewC(t-i) * e^{\frac{t-i}{\tau_{RewC}}} + \beta_0 \quad [\mathbf{eq.8}]$$

where $P_L(t)$ is the probability of choosing left on trial t, $RewC(t-i)$ is the rewarded choice on past trial $t-i$ (1 if rewarded on the left, -1 if rewarded on the right, 0 otherwise). Up to 15 past trials were considered for this model (N=15), unless otherwise noted. β_{RewC} is the linear regression weight on the kernel, β_0 is the constant bias term, and τ_{RewC} is the time-constant of the exponential.

Similarly, the hyperbolic model was defined as.

$$logit\big(P_L(t)\big) = \beta_{RewC} * \sum_{i=1}^{N} RewC(t-i) * \frac{1}{1+\frac{i-1}{\tau_{RewC}}} + \beta_0 \qquad [\mathbf{eq.9}]$$

The only difference from the exponential model is the form of the decay function, with the time-constant τ_{RewC} of the hyperbolic. For both models, τ_{RewC} was constrained to be greater than 0. These models were fit to the choice patterns of 74 sessions of expert mouse behavior in python using SciPy *minimize* function, with search algorithm L-BFGS-B, to perform maximum likelihood estimation

To compare the performance of each model, each session was divided into ten equal sets of trials, nine of which were used to estimate the exponential and hyperbolic models. At each iteration the log of the likelihood for the held-out trials was compared as $loglik_{hyp} - loglik_{exp}$. This was iterated across each held-out test set, and the difference in loglikelihood taken as the mean across all iterations, yielding cross-validated (CV) loglikelihood.

3.4.12 Exponential and Hyperbolic behavioral models for optogenetic analysis

The exponential and hyperbolic behavioral models under optogenetic inactivation were fit with the logistic regression analysis, by selecting only the inactivation or pseudo-inactivation trials from the inactivation or control sessions, and concatenating the trials from

all sessions for a given mouse, while using a mixed-effects model to account for across-session variabilities. The concatenated trials were fit with [eq. 8] and [eq. 9] separately for the control condition or the inactivation condition, with a random intercept for each session as in [eq. 4]. Given that inactivation trials occurred on only 15% of trials, the number of trials held-out to calculate the CV-loglikelihood could be very small, sometimes <10 trials. To increase the consistency of the loglikelihood across animals, we normalized loglikelihood by dividing by the number of trials in the test-set to yield CV-loglikelihood per trial.

3.4.13 Exponential and Hyperbolic cell models

To identify the neurons modulated by rewarded-choice history, we averaged the neural activity during the first 2 s of the ready-period, during which the mouse was withholding licking. Then we estimated the influence of the most recent rewarded-choice trial as:

$$A(t) = \beta_{RewC} * RewC(t-1) + \beta_C C(t) + \beta_0 \quad \textbf{[eq. 10]}$$

Where $A(t)$ is the neural activity at trial t, $RewC(t-1)$ is the rewarded choice on immediately preceding trial (1 if rewarded on the left, -1 if rewarded on the right, 0 otherwise), and $C(t)$ is the choice that the animal will make on the current trial (1 if left, -1 if right) to regress out anticipatory movement-related activity. β_{RewC} is the linear regression weight on the rewarded-choice history, β_C is the weight for the upcoming choice, and β_0 is baseline offset. Neurons with a p-value < 0.05 for β_{RewC} were considered modulated by past rewarded-choice.

Among these modulated neurons, we then calculated whether they were more exponential or hyperbolic in their history integration, in an analogous method to the above behavior models [eq. 8 & 9]. Specifically, neural activity was fit by the exponential model:

$$A(t) = \beta_{RewC} * \sum_{i=1}^{N} RewC(t-i) * e^{\frac{t-i}{\tau_{RewC}}} + \beta_C C(t) + \beta_0 \quad [\mathbf{eq.11}]$$

and the hyperbolic model:

$$A(t) = \beta_{RewC} * \sum_{i=1}^{N} RewC(t-i) * \frac{1}{1+\frac{i-1}{\tau_{RewC}}} + \beta_C C(t) + \beta_0 \quad [\mathbf{eq.12}]$$

Performance of the two models was compared for each neuron as the loglikelihood of the 10-fold cross-validated test set, and the difference taken as $loglik_{hyp} - loglik_{exp}$ for each iteration.

For those cells that were consistently modulated by past-choice [eq. 10] we further estimated the exponential time-constant across all trials in the session, and from two non-overlapping halves of the session. Cells were considered exponentially-modulated if the p–value for β_{RewC} was <0.05 in both the first half and second half of the session, though no constraint was imposed that either β_{RewC} or τ_{RewC} be consistent between halves. The full session τ_{RewC} was used for all analysis except for Figure 3.3 D-E, where the two independently estimated τ_{RewC} were compared as a metric of the stability of neural encoding and model estimation.

3.4.14 Quasi-hyperbolic behavioral model

The quasi-hyperbolic model is defined as a set of weighted exponential functions summing together to yield a probability of choosing left or right. From the observed distributions of τ_{RewC} of exponential neurons in each cortical area, we randomly drew between 1 and 15 values of τ and fit the weighting on each exponential kernel necessary to best describe the behavior, following the equation:

$$logit\big(P_L(t)\big) = \sum_{m=i}^{M} \beta_m * \sum_{i=1}^{N} RewC(t-i) * e^{\frac{t-i}{\tau_m}} + \beta_0 \quad [\mathbf{eq.13}]$$

where $P_L(t)$ is the probability of choosing left on trial t, $RewC(t-i)$ is the rewarded choice on past trial $t-i$ (1 if rewarded on the left, -1 if rewarded on the right, 0 otherwise), β_m is the linear coefficient corresponding to the exponential kernel with time-constant τ_m, and β_0 is the constant bias. The number of past trials considered was $N = 15$. Each set of τ was fit to each of the 74 behavior sessions to yield the 10-fold cross-validated loglikelihood, similarly to the exponential and hyperbolic behavior models [eq. 8 & 9].

3.4.15 Statistical tests

Paired comparisons made with two-tailed paired t-test if data passed normality by Lilliefors test, otherwise with Wilcoxon signed-rank, as noted. Unpaired comparisons with two-tailed independent t-test if data passed normality by Lilliefors test, otherwise with Wilcoxon rank sum, as noted. Reported p-values were false discovery rate (FDR) corrected with Benjamini-Hochberg method as appropriate for multiple comparisons. All loglikelihood measurements were assessed with 10-fold cross validation. In some analyses where the number of trials in the test set was very small, and therefore the likelihood could be noisy across folds, the loglikelihood was normalized by dividing it by the number of trials in the test set, producing loglikelihood per trial. The non-parametric test of distribution medians was bootstrapped 100,000 times, with an equal number of samples drawn from each cortical area, with replacement. Mixed effects modeling included random slope and intercept for session identity or mouse identity, and was estimated with the R library in Python via the lme4 package. Two-way repeated measures ANOVA was performed in Python with the package

pingouin. All other statistical tests were performed in Python with either SciPy or Statsmodels.

3.5 Acknowledgements

Chapter 3, in full, is material currently being prepared for submission for publication. Danskin B, Hattori R, Zhang EY, Aoi M, Komiyama T. Diverse behavioral timescales encoded in retrosplenial cortex explain hyperbolic behavior. The dissertation author was the primary author of this material.

References

Alexander, William H., and Joshua W. Brown. 2010. “Hyperbolically Discounted Temporal Difference Learning.” *Neural Computation* 22 (6): 1511–27. https://doi.org/10.1162/neco.2010.08-09-1080.

Aparicio, Carlos F., and William M. Baum. 2009. “Dynamics of Choice: Relative Rate and Amount Affect Local Preference at Three Different Time Scales.” *Journal of the Experimental Analysis of Behavior* 91 (3): 293–317. https://doi.org/10.1901/jeab.2009.91-293.

Behrens, Timothy E J, Mark W Woolrich, Mark E Walton, and Matthew F S Rushworth. 2007. “Learning the Value of Information in an Uncertain World.” *Nature Neuroscience* 10 (9): 1214–21. https://doi.org/10.1038/nn1954.

Bernacchia, Alberto, Hyojung Seo, Daeyeol Lee, and Xiao-Jing Wang. 2011. “A Reservoir of Time Constants for Memory Traces in Cortical Neurons - Supplement.” *Nature Neuroscience* 14 (3): 366–72. https://doi.org/10.1038/nn.2752.

Brennan, Ellen K.W., Shyam Kumar Sudhakar, Izabela Jedrasiak-Cape, Tibin T. John, and Omar J. Ahmed. 2020. “Hyperexcitable Neurons Enable Precise and Persistent Information Encoding in the Superficial Retrosplenial Cortex.” *Cell Reports* 30 (5): 1598-1612.e8. https://doi.org/10.1016/j.celrep.2019.12.093.

Brody, C. D. 2003. “Timing and Neural Encoding of Somatosensory Parametric Working Memory in Macaque Prefrontal Cortex.” *Cerebral Cortex* 13 (11): 1196–1207. https://doi.org/10.1093/cercor/bhg100.

Cavanagh, Sean E, Joni D Wallis, Steven W Kennerley, and Laurence T Hunt. 2016. “Autocorrelation Structure at Rest Predicts Value Correlates of Single Neurons during Reward-Guided Choice.” *ELife* 5 (October): e18937. https://doi.org/10.7554/eLife.18937.

Chaudhuri, Rishidev, Kenneth Knoblauch, Marie-Alice Gariel, Henry Kennedy, and Xiao-Jing Wang. 2015. “A Large-Scale Circuit Mechanism for Hierarchical Dynamical Processing in the Primate Cortex.” *Neuron* 88 (2): 419–31. https://doi.org/10.1016/j.neuron.2015.09.008.

Cohen, Jeremiah Y, Mackenzie W Amoroso, and Naoshige Uchida. 2015. “Serotonergic Neurons Signal Reward and Punishment on Multiple Timescales.” *ELife* 4 (February): e06346. https://doi.org/10.7554/eLife.06346.

Corrado, Greg S., Leo P. Sugrue, H. Sebastian Seung, and William T. Newsome. 2005. “Linear-Nonlinear-Poisson Models of Primate Choice Dynamics.” *Journal of the Experimental Analysis of Behavior* 84 (3): 581–617. https://doi.org/10.1901/jeab.2005.23-05.

Dabney, Will, Zeb Kurth-Nelson, Naoshige Uchida, Clara Kwon Starkweather, Demis Hassabis, Rémi Munos, and Matthew Botvinick. 2020. “A Distributional Code for Value in

Dopamine-Based Reinforcement Learning." *Nature* 577 (7792): 671–75. https://doi.org/10.1038/s41586-019-1924-6.

Daw, Nathaniel D, Yael Niv, and Peter Dayan. 2005. "Uncertainty-Based Competition between Prefrontal and Dorsolateral Striatal Systems for Behavioral Control." *Nature Neuroscience* 8 (12): 1704–11. https://doi.org/10.1038/nn1560.

Daw, Nathaniel D., John P. O'Doherty, Peter Dayan, Ben Seymour, and Raymond J. Dolan. 2006. "Cortical Substrates for Exploratory Decisions in Humans." *Nature* 441 (7095): 876–79. https://doi.org/10.1038/nature04766.

Daw, Nathaniel D., Samuel J. Gershman, Ben Seymour, Peter Dayan, and Raymond J. Dolan. 2011. "Model-Based Influences on Humans' Choices and Striatal Prediction Errors." *Neuron* 69 (6): 1204–15. https://doi.org/10.1016/j.neuron.2011.02.027.

Dayan, Peter, Sham Kakade, and P. Read Montague. 2000. "Learning and Selective Attention." *Nature Neuroscience* 3 (S11): 1218–23. https://doi.org/10.1038/81504.

Dhawale, Ashesh K, Akari Hagiwara, Upinder S Bhalla, Venkatesh N Murthy, and Dinu F Albeanu. 2010. "Non-Redundant Odor Coding by Sister Mitral Cells Revealed by Light Addressable Glomeruli in the Mouse." *Nature Neuroscience* 13 (11): 1404–12. https://doi.org/10.1038/nn.2673.

Doya, Kenji. 2002. "Metalearning and Neuromodulation." *Neural Networks* 15 (4–6): 495–506. https://doi.org/10.1016/S0893-6080(02)00044-8.

Fedus, William, Carles Gelada, Yoshua Bengio, Marc G. Bellemare, and Hugo Larochelle. 2019. "Hyperbolic Discounting and Learning over Multiple Horizons." *ArXiv:1902.06865 [Cs, Stat]*, February. http://arxiv.org/abs/1902.06865.

Frederick, Shane, and George Loewenstein. 2002. "Time Discounting and Time Preference: A Critical Review," 52.

Friedrich, Johannes, Pengcheng Zhou, and Liam Paninski. 2017. "Fast Online Deconvolution of Calcium Imaging Data." Edited by Joshua Vogelstein. *PLOS Computational Biology* 13 (3): e1005423. https://doi.org/10.1371/journal.pcbi.1005423.

Goldman, Mark S. 2017. "Memory without Feedback in a Neural Network." *Neuron* 93 (3): 715. https://doi.org/10.1016/j.neuron.2017.01.007.

Grossman, Cooper D., Bilal A. Bari, and Jeremiah Y. Cohen. 2022. "Serotonin Neurons Modulate Learning Rate through Uncertainty." *Current Biology* 32 (3): 586-599.e7. https://doi.org/10.1016/j.cub.2021.12.006.

Haddad, Rafi, Anne Lanjuin, Linda Madisen, Hongkui Zeng, Venkatesh N Murthy, and Naoshige Uchida. 2013. “Olfactory Cortical Neurons Read out a Relative Time Code in the Olfactory Bulb.” *Nature Neuroscience* 16 (7): 949–57. https://doi.org/10.1038/nn.3407.

Haith, A. M., T. R. Reppert, and R. Shadmehr. 2012. “Evidence for Hyperbolic Temporal Discounting of Reward in Control of Movements.” *Journal of Neuroscience* 32 (34): 11727–36. https://doi.org/10.1523/JNEUROSCI.0424-12.2012.

Harvey, Christopher D., Philip Coen, and David W. Tank. 2012. “Choice-Specific Sequences in Parietal Cortex during a Virtual-Navigation Decision Task.” *Nature* 484 (7392): 62–68. https://doi.org/10.1038/nature10918.

Hattori, Ryoma, Bethanny Danskin, Zeljana Babic, Nicole Mlynaryk, and Takaki Komiyama. 2019. “Area-Specificity and Plasticity of History-Dependent Value Coding During Learning.” *Cell* 177 (7): 1858-1872.e15. https://doi.org/10.1016/j.cell.2019.04.027.

Hattori, Ryoma, and Takaki Komiyama. 2022a. “Context-Dependent Persistency as a Coding Mechanism for Robust and Widely Distributed Value Coding.” *Neuron* 110 (3): 502-515.e11. https://doi.org/10.1016/j.neuron.2021.11.001.

Hattori, Ryoma, and Takaki Komiyama. 2022b. “PatchWarp: Corrections of Non-Uniform Image Distortions in Two-Photon Calcium Imaging Data by Patchwork Affine Transformations.” *Cell Reports Methods*, April, 100205. https://doi.org/10.1016/j.crmeth.2022.100205.

Hattori, Ryoma, and Takaki Komiyama. 2022c. “Longitudinal Two-Photon Calcium Imaging with Ultra-Large Cranial Window for Head-Fixed Mice.” *STAR Protocols* 3 (2): 101343. https://doi.org/10.1016/j.xpro.2022.101343.

Hunt, Laurence T., and Benjamin Y. Hayden. 2017. “A Distributed, Hierarchical and Recurrent Framework for Reward-Based Choice.” *Nature Reviews Neuroscience* 18 (3): 172–82. https://doi.org/10.1038/nrn.2017.7.

Iigaya, Kiyohito, Yashar Ahmadian, Leo P. Sugrue, Greg S. Corrado, Yonatan Loewenstein, William T. Newsome, and Stefano Fusi. 2019. “Deviation from the Matching Law Reflects an Optimal Strategy Involving Learning over Multiple Timescales.” *Nature Communications* 10 (1): 1466. https://doi.org/10.1038/s41467-019-09388-3.

Ito, M., and K. Doya. 2009. “Validation of Decision-Making Models and Analysis of Decision Variables in the Rat Basal Ganglia.” *Journal of Neuroscience* 29 (31): 9861–74. https://doi.org/10.1523/JNEUROSCI.6157-08.2009.

Kennerley, Steven W, Mark E Walton, Timothy E J Behrens, Mark J Buckley, and Matthew F S Rushworth. 2006. “Optimal Decision Making and the Anterior Cingulate Cortex.” *Nature Neuroscience* 9 (7): 940–47. https://doi.org/10.1038/nn1724.

Kim, HyungGoo R., Athar N. Malik, John G. Mikhael, Pol Bech, Iku Tsutsui-Kimura, Fangmiao Sun, Yajun Zhang, et al. 2020. “A Unified Framework for Dopamine Signals across Timescales.” *Cell* 183 (6): 1600-1616.e25. https://doi.org/10.1016/j.cell.2020.11.013.

Kurth-Nelson, Zeb, and A. David Redish. 2009. “Temporal-Difference Reinforcement Learning with Distributed Representations.” Edited by Olaf Sporns. *PLoS ONE* 4 (10): e7362. https://doi.org/10.1371/journal.pone.0007362.

Lau, Brian, and Paul W. Glimcher. 2005. “Dynamic Response-by-Response Models of Matching Behavior in Rhesus Monkeys.” *Journal of the Experimental Analysis of Behavior* 84 (3): 555–79. https://doi.org/10.1901/jeab.2005.110-04.

Lowet, Adam S., Qiao Zheng, Sara Matias, Jan Drugowitsch, and Naoshige Uchida. 2020. “Distributional Reinforcement Learning in the Brain.” *Trends in Neurosciences* 43 (12): 980–97. https://doi.org/10.1016/j.tins.2020.09.004.

Massi, Bart, Christopher H. Donahue, and Daeyeol Lee. 2018. “Volatility Facilitates Value Updating in the Prefrontal Cortex.” *Neuron* 99 (3): 598-608.e4. https://doi.org/10.1016/j.neuron.2018.06.033.

Matias, Sara, Eran Lottem, Guillaume P Dugué, and Zachary F Mainen. 2017. “Activity Patterns of Serotonin Neurons Underlying Cognitive Flexibility.” *ELife* 6 (March): e20552. https://doi.org/10.7554/eLife.20552.

Meder, David, Nils Kolling, Lennart Verhagen, Marco K. Wittmann, Jacqueline Scholl, Kristoffer H. Madsen, Oliver J. Hulme, Timothy E.J. Behrens, and Matthew F.S. Rushworth. 2017. “Simultaneous Representation of a Spectrum of Dynamically Changing Value Estimates during Decision Making.” *Nature Communications* 8 (1): 1942. https://doi.org/10.1038/s41467-017-02169-w.

Mitani, Akinori, and Takaki Komiyama. 2018. “Real-Time Processing of Two-Photon Calcium Imaging Data Including Lateral Motion Artifact Correction.” *Frontiers in Neuroinformatics* 12 (December): 98. https://doi.org/10.3389/fninf.2018.00098.

Miyazaki, Katsuhiko, Kayoko W. Miyazaki, Gaston Sivori, Akihiro Yamanaka, Kenji F. Tanaka, and Kenji Doya. 2020. “Serotonergic Projections to the Orbitofrontal and Medial Prefrontal Cortices Differentially Modulate Waiting for Future Rewards.” *Science Advances* 6 (48): eabc7246. https://doi.org/10.1126/sciadv.abc7246.

Miyazaki, Kayoko W., Katsuhiko Miyazaki, Kenji F. Tanaka, Akihiro Yamanaka, Aki Takahashi, Sawako Tabuchi, and Kenji Doya. 2014. “Optogenetic Activation of Dorsal Raphe Serotonin Neurons Enhances Patience for Future Rewards.” *Current Biology* 24 (17): 2033–40. https://doi.org/10.1016/j.cub.2014.07.041.

Murray, John D, Alberto Bernacchia, David J Freedman, Ranulfo Romo, Jonathan D Wallis, Xinying Cai, Camillo Padoa-Schioppa, et al. 2014. “A Hierarchy of Intrinsic Timescales

across Primate Cortex." *Nature Neuroscience* 17 (12): 1661–63. https://doi.org/10.1038/nn.3862.

Pachitariu, Marius, Carsen Stringer, Mario Dipoppa, Sylvia Schröder, L. Federico Rossi, Henry Dalgleish, Matteo Carandini, and Kenneth D. Harris. 2016. "Suite2p: Beyond 10,000 Neurons with Standard Two-Photon Microscopy." Preprint. Neuroscience. https://doi.org/10.1101/061507.

Pachitariu, Marius, Carsen Stringer, and Kenneth D. Harris. 2018. "Robustness of Spike Deconvolution for Neuronal Calcium Imaging." *The Journal of Neuroscience* 38 (37): 7976–85. https://doi.org/10.1523/JNEUROSCI.3339-17.2018.

Rescorla, R., and A. Wagner. 1972. "A Theory of Pavlovian Conditioning : Variations in the Effectiveness of Reinforcement and Nonreinforcement." In *Classical Conditioning II: Current Research and Theory*, 64–99. Appleton-Century-Crofts.

Runyan, Caroline A., Eugenio Piasini, Stefano Panzeri, and Christopher D. Harvey. 2017. "Distinct Timescales of Population Coding across Cortex." *Nature* 548 (7665): 92–96. https://doi.org/10.1038/nature23020.

Scott, Benjamin B., Christine M. Constantinople, Athena Akrami, Timothy D. Hanks, Carlos D. Brody, and David W. Tank. 2017. "Fronto-Parietal Cortical Circuits Encode Accumulated Evidence with a Diversity of Timescales." *Neuron* 95 (2): 385-398.e5. https://doi.org/10.1016/j.neuron.2017.06.013.

Serences, John T. 2008. "Value-Based Modulations in Human Visual Cortex." *Neuron* 60 (6): 1169–81. https://doi.org/10.1016/j.neuron.2008.10.051.

Spitmaan, Mehran, Hyojung Seo, Daeyeol Lee, and Alireza Soltani. 2020. "Multiple Timescales of Neural Dynamics and Integration of Task-Relevant Signals across Cortex." *Proceedings of the National Academy of Sciences* 117 (36): 22522–31. https://doi.org/10.1073/pnas.2005993117.

Sugrue, L. P. 2004. "Matching Behavior and the Representation of Value in the Parietal Cortex." *Science* 304 (5678): 1782–87. https://doi.org/10.1126/science.1094765.

Sutton, Richard S., and Andrew G. Barto. 1998. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Tiganj, Zoran, Michael E. Hasselmo, and Marc W. Howard. 2015. "A Simple Biophysically Plausible Model for Long Time Constants in Single Neurons." *Hippocampus* 25 (1): 27–37. https://doi.org/10.1002/hipo.22347.

Wilson, Robert C., Matthew R. Nassar, and Joshua I. Gold. 2013. "A Mixture of Delta-Rules Approximation to Bayesian Inference in Change-Point Problems." Edited by Tim Behrens. *PLoS Computational Biology* 9 (7): e1003150. https://doi.org/10.1371/journal.pcbi.1003150.

Yu, Angela J., and Peter Dayan. 2005. “Uncertainty, Neuromodulation, and Attention.” *Neuron* 46 (4): 681–92. https://doi.org/10.1016/j.neuron.2005.04.026.

Experimental work in the behaving mouse indicates that RSC encodes history-dependent value with a diversity of timescales, and that RSC is acutely necessary in using this information for decision-making. The temporal information encoded in RSC develops as a function of task learning (Hattori et al. 2019), is persistently available to the animal (Hattori et al. 2019; Hattori and Komiyama 2022), and the distribution of timescales can explain the hyperbolic behavioral strategy of the animal. This accumulated evidence demonstrates that RSC provides a flexible and adaptable neural substrate for history-based decision-making.

4.1 Expansion of the reinforcement learning framework.

As discussed in Chapters 1 and 2, reinforcement learning theory provides a computational grounding from which to describe behavior in a quantitative way. However, to quote George Box, all models are wrong but some models are useful. The RL models describe fundamental and proven aspects of reward-choice association, but may miss certain aspects of nuanced decision-making. Specifically, Chapter 3 of this dissertation examines one such critical 'miss' of standard RL models: hyperbolic decay of past experience as well as prospective hyperbolic discounting. The dissertation author's proposed model of distributed temporal integrators, each of which might individually act as RL agents following Q-learning rules, forms a bridge between theories of reinforcement learning and observed behavioral phenomena. Supporting this notion is the representation of diverse timescales of history integration in the cortex, and in particular in RSC. This encoding produces a 'reservoir' of temporal information persistently available to the animal.

Notably, multiple groups working across different animal models, theoretical models, and brain areas have converged to this idea of distributed integration of reward-history related

information (Bernacchia et al. 2011; Spitmaan et al. 2020; Dabney et al. 2020). Of particular relevance are models of distributional reinforcement learning (Dabney et al. 2020; Lowet et al. 2020), which describe decision as made by an agent with a distributed representation of predictions about the future reward space. One hypothesis of the model is that neurons encoding different estimates of value exist in separate, independently updating subnetworks of neurons. While previous research has identified such neurons in subcortical structures, namely the ventral tegmental area (Dabney et al. 2020), this author proposes that the neurons in RSC cortex may act as the cortical partners in these networks. Cortical-subcortical reverberations might then take advantage of the recurrent connectivity in RSC to maintain behaviorally-relevant information over seconds to minutes during decision-making.

4.2 Properties of retrosplenial cortex.

Retrosplenial cortex has a number of properties which make it ideally suited to be a reward-history associational area, principally that it combines information from multiple modalities and interfaces with the hippocampal working-memory system. RSC is known to integrate spatial and contextual information, and is densely connected with visual cortex, posterior parietal cortex, prefrontal cortex, and entorhinal cortex (van Groen and Michael Wyss 1990; van Groen and Wyss 1992; Vann, Aggleton, and Maguire 2009).

In addition to dense projection patterns, RSC is also notable for its population of highly excitable, low-rheobase cells and strong local inhibition (Brennan et al. 2020). The combination of strong feedforward excitation and inhibition yields a neural network capable of producing sustained and specific responses to outside input (Brennan et al. 2020). The pattern of population-space activity of RSC neurons has been found to be particularly

persistent, indicating the capacity for sustained encoding of behavior-relevant information (Hattori and Komiyama 2022).

RSC neurons exhibit great heterogeneity in their encoding of task-variables, crossing modalities and often multiplexing different types of information in the same neurons (Hattori et al. 2019; Sun et al. 2021). Perturbation experiments including acute inactivation (Hattori et al. 2019), chemogenetic inhibition (Sun et al. 2021), and lesion (Vann, Aggleton, and Maguire 2009) indicate that RSC is necessary for both initial learning and ongoing choice-outcome associations, but not necessarily for retrieval of previously learned associations. This is consistent with the dichotomy observed in this dissertation between acute inactivation of RSC, which produced deficits in ongoing incorporation of information, and lesion of RSC, which produced no such deficits. In the chronic absence of RSC the animal adapted to use information that is redundantly encoded in the brain, but natively RSC appears to be in the default path for labile value update.

4.3 A dynamic reservoir of temporal information.

Having a distributed representation of information across a range of temporal horizons provides a mechanism by which an animal may maintain information efficiently, but alter readout rapidly. Such rapid adaptation is beneficial when environments are dynamic, shifting between periods stability to volatility (Dayan, Kakade, and Montague 2000; Daw et al. 2006; Daw, Niv, and Dayan 2005; Daw et al. 2011; Kennerley et al. 2006; T. E. J. Behrens et al. 2007; Meder et al. 2017; Massi, Donahue, and Lee 2018; Bernacchia et al. 2011; Iigaya et al. 2019). An alternate explanation for the hyperbolic-like integration, one that is not mutually exclusive to the model detailed in chapter 3, is that animals shift between multiple discrete

strategies throughout one analytic period (i.e. one session), and that these strategies may integrate information with different timescales.

There is evidence that animals change temporal integration strategies rapidly, both in cases where this is adaptive for the animal (T. E. J. Behrens et al. 2007; Nassar et al. 2010; Wilson, Nassar, and Gold 2013; Lee, Shimojo, and O'Doherty 2014; Massi, Donahue, and Lee 2018; D. Kim et al. 2019; Akam et al. 2021), and cases where this is an incidental behavior that is nevertheless quantifiable (Balcarras et al. 2016; Beron et al. 2021; Ashwood et al. 2022). Shifting between temporal integration strategies on a trial-to-trial basis can produce the appearance of heavy-tailed history weighting (Iigaya et al. 2019), whether the strategy employed on any given trial is to use history weighted with exponential or hyperbolic decay. Further modeling analysis with greater temporal specificity may be necessary to evaluate whether the choice on a single trial is exponentially or hyperbolically weighted, but the strategy across trials clearly incorporates history information across multiple temporal windows.

Rewarded-choice history information is simultaneously encoded with a diversity of timescales across cells. What remains an ongoing area of investigation is how that combination of different timescales might change to enhance shorter or longer integration as the animal shifts its behavioral strategy. Here the author poses two distinct hypotheses for how a change in the behavior could be derived from the distributed temporal information available in RSC. First, that individual neurons alter their time-constants of integration across time, according to environment demands, perhaps as a consequence of population-level shifts mediated by neuromodulation by dopamine (Doya 2002; Ito and Doya 2009; H. R. Kim et al. 2020), noradrenaline (Doya 2002; Yu and Dayan 2005), or serotonin (Doya 2002; Cohen,

Amoroso, and Uchida 2015; Matias et al. 2017; Grossman, Bari, and Cohen 2022). Second, that an individual neuron has a consistent time-constant, but the gain of different neurons across the population is differentially modulated such that the observed distribution of temporal integrators changes. This population-space shift would be readout by downstream circuits, and produce behavior with shorter or longer integration. Alternately, this selective processing of temporal information may occur solely in a downstream area, and RSC provides a stable reservoir of information even during changing behavior. Future work that examines a longitudinal recording of cells across sessions, or across a more volatile environment, may be able to disambiguate these possibilities.

4.4 Conclusion

Retrosplenial cortex is a highly interconnected cortical area that plays a critical role in decision-making and associational learning. In particular, the encoding of diverse temporal information in RSC and the persistency of value-coding suggest RSC acts as a reservoir of behaviorally-relevant information. Damage to RSC has deleterious effects on cognition and decision-making (Maguire 2001; J. H. Kim et al. 2007), and loss of RSC neuron mass is associated with early-stage symptoms in Alzheimer's dementia (Minoshima et al. 1997; Nestor et al. 2003; Pengas et al. 2010; Tan et al. 2013). These clinical observations are consistent with the role of RSC as a central hub in decision-making, and as a neural substrate primed to integrate different modes of information across both short and long periods of time.

References

Akam, Thomas, Ines Rodrigues-Vaz, Ivo Marcelo, Xiangyu Zhang, Michael Pereira, Rodrigo Freire Oliveira, Peter Dayan, and Rui M. Costa. 2021. "The Anterior Cingulate Cortex Predicts Future States to Mediate Model-Based Action Selection." *Neuron* 109 (1): 149-163.e7. https://doi.org/10.1016/j.neuron.2020.10.013.

Ashwood, Zoe C., Nicholas A. Roy, Iris R. Stone, The International Brain Laboratory, Anne E. Urai, Anne K. Churchland, Alexandre Pouget, and Jonathan W. Pillow. 2022. "Mice Alternate between Discrete Strategies during Perceptual Decision-Making." *Nature Neuroscience* 25 (2): 201–12. https://doi.org/10.1038/s41593-021-01007-z.

Balcarras, Matthew, Salva Ardid, Daniel Kaping, Stefan Everling, and Thilo Womelsdorf. 2016. "Attentional Selection Can Be Predicted by Reinforcement Learning of Task-Relevant Stimulus Features Weighted by Value-Independent Stickiness." *Journal of Cognitive Neuroscience* 28 (2): 333–49. https://doi.org/10.1162/jocn_a_00894.

Behrens, Timothy E J, Mark W Woolrich, Mark E Walton, and Matthew F S Rushworth. 2007. "Learning the Value of Information in an Uncertain World." *Nature Neuroscience* 10 (9): 1214–21. https://doi.org/10.1038/nn1954.

Bernacchia, Alberto, Hyojung Seo, Daeyeol Lee, and Xiao-Jing Wang. 2011. "A Reservoir of Time Constants for Memory Traces in Cortical Neurons - Supplement." *Nature Neuroscience* 14 (3): 366–72. https://doi.org/10.1038/nn.2752.

Beron, Celia, Shay Neufeld, Scott Linderman, and Bernardo Sabatini. 2021. "Efficient and Stochastic Mouse Action Switching during Probabilistic Decision Making." Preprint. Neuroscience. https://doi.org/10.1101/2021.05.13.444094.

Brennan, Ellen K.W., Shyam Kumar Sudhakar, Izabela Jedrasiak-Cape, Tibin T. John, and Omar J. Ahmed. 2020. "Hyperexcitable Neurons Enable Precise and Persistent Information Encoding in the Superficial Retrosplenial Cortex." *Cell Reports* 30 (5): 1598-1612.e8. https://doi.org/10.1016/j.celrep.2019.12.093.

Cohen, Jeremiah Y, Mackenzie W Amoroso, and Naoshige Uchida. 2015. "Serotonergic Neurons Signal Reward and Punishment on Multiple Timescales." *ELife* 4 (February): e06346. https://doi.org/10.7554/eLife.06346.

Dabney, Will, Zeb Kurth-Nelson, Naoshige Uchida, Clara Kwon Starkweather, Demis Hassabis, Rémi Munos, and Matthew Botvinick. 2020. "A Distributional Code for Value in Dopamine-Based Reinforcement Learning." *Nature* 577 (7792): 671–75. https://doi.org/10.1038/s41586-019-1924-6.

Daw, Nathaniel D, Yael Niv, and Peter Dayan. 2005. "Uncertainty-Based Competition between Prefrontal and Dorsolateral Striatal Systems for Behavioral Control." *Nature Neuroscience* 8 (12): 1704–11. https://doi.org/10.1038/nn1560.

Daw, Nathaniel D., John P. O'Doherty, Peter Dayan, Ben Seymour, and Raymond J. Dolan. 2006. "Cortical Substrates for Exploratory Decisions in Humans." *Nature* 441 (7095): 876–79. https://doi.org/10.1038/nature04766.

Daw, Nathaniel D., Samuel J. Gershman, Ben Seymour, Peter Dayan, and Raymond J. Dolan. 2011. "Model-Based Influences on Humans' Choices and Striatal Prediction Errors." *Neuron* 69 (6): 1204–15. https://doi.org/10.1016/j.neuron.2011.02.027.

Dayan, Peter, Sham Kakade, and P. Read Montague. 2000. "Learning and Selective Attention." *Nature Neuroscience* 3 (S11): 1218–23. https://doi.org/10.1038/81504.

Doya, Kenji. 2002. "Metalearning and Neuromodulation." *Neural Networks* 15 (4–6): 495–506. https://doi.org/10.1016/S0893-6080(02)00044-8.

Groen, Thomas van, and J. Michael Wyss. 1990. "Connections of the Retrosplenial Granular a Cortex in the Rat." *The Journal of Comparative Neurology* 300 (4): 593–606. https://doi.org/10.1002/cne.903000412.

Groen, Thomas van, and J. Michael Wyss. 1992. "Connections of the Retrosplenial Dysgranular Cortex in the Rat." *The Journal of Comparative Neurology* 315 (2): 200–216. https://doi.org/10.1002/cne.903150207.

Grossman, Cooper D., Bilal A. Bari, and Jeremiah Y. Cohen. 2022. "Serotonin Neurons Modulate Learning Rate through Uncertainty." *Current Biology* 32 (3): 586-599.e7. https://doi.org/10.1016/j.cub.2021.12.006.

Hattori, Ryoma, Bethanny Danskin, Zeljana Babic, Nicole Mlynaryk, and Takaki Komiyama. 2019. "Area-Specificity and Plasticity of History-Dependent Value Coding During Learning." *Cell* 177 (7): 1858-1872.e15. https://doi.org/10.1016/j.cell.2019.04.027.

Hattori, Ryoma, and Takaki Komiyama. 2022. "Context-Dependent Persistency as a Coding Mechanism for Robust and Widely Distributed Value Coding." *Neuron* 110 (3): 502-515.e11. https://doi.org/10.1016/j.neuron.2021.11.001.

Iigaya, Kiyohito, Yashar Ahmadian, Leo P. Sugrue, Greg S. Corrado, Yonatan Loewenstein, William T. Newsome, and Stefano Fusi. 2019. "Deviation from the Matching Law Reflects an Optimal Strategy Involving Learning over Multiple Timescales." *Nature Communications* 10 (1): 1466. https://doi.org/10.1038/s41467-019-09388-3.

Ito, M., and K. Doya. 2009. "Validation of Decision-Making Models and Analysis of Decision Variables in the Rat Basal Ganglia." *Journal of Neuroscience* 29 (31): 9861–74. https://doi.org/10.1523/JNEUROSCI.6157-08.2009.

Kennerley, Steven W, Mark E Walton, Timothy E J Behrens, Mark J Buckley, and Matthew F S Rushworth. 2006. "Optimal Decision Making and the Anterior Cingulate Cortex." *Nature Neuroscience* 9 (7): 940–47. https://doi.org/10.1038/nn1724.

Kim, Dongjae, Geon Yeong Park, John P. O′Doherty, and Sang Wan Lee. 2019. “Task Complexity Interacts with State-Space Uncertainty in the Arbitration between Model-Based and Model-Free Learning.” *Nature Communications* 10 (1): 5738. https://doi.org/10.1038/s41467-019-13632-1.

Kim, HyungGoo R., Athar N. Malik, John G. Mikhael, Pol Bech, Iku Tsutsui-Kimura, Fangmiao Sun, Yajun Zhang, et al. 2020. “A Unified Framework for Dopamine Signals across Timescales.” *Cell* 183 (6): 1600-1616.e25. https://doi.org/10.1016/j.cell.2020.11.013.

Kim, Jong Hun, Kwang-Yeol Park, Sang Won Seo, Duk L. Na, Chin-Sang Chung, Kwang Ho Lee, and Gyeong-Moon Kim. 2007. “Reversible Verbal and Visual Memory Deficits after Left Retrosplenial Infarction.” *Journal of Clinical Neurology* 3 (1): 62. https://doi.org/10.3988/jcn.2007.3.1.62.

Lee, Sang Wan, Shinsuke Shimojo, and John P. O’Doherty. 2014. “Neural Computations Underlying Arbitration between Model-Based and Model-Free Learning.” *Neuron* 81 (3): 687–99. https://doi.org/10.1016/j.neuron.2013.11.028.

Lowet, Adam S., Qiao Zheng, Sara Matias, Jan Drugowitsch, and Naoshige Uchida. 2020. “Distributional Reinforcement Learning in the Brain.” *Trends in Neurosciences* 43 (12): 980–97. https://doi.org/10.1016/j.tins.2020.09.004.

Maguire, Eleanor A. 2001. “The Retrosplenial Contribution to Human Navigation: A Review of Lesion and Neuroimaging Findings.” *Scandinavian Journal of Psychology* 42 (3): 225–38.

Massi, Bart, Christopher H. Donahue, and Daeyeol Lee. 2018. “Volatility Facilitates Value Updating in the Prefrontal Cortex.” *Neuron* 99 (3): 598-608.e4. https://doi.org/10.1016/j.neuron.2018.06.033.

Matias, Sara, Eran Lottem, Guillaume P Dugué, and Zachary F Mainen. 2017. “Activity Patterns of Serotonin Neurons Underlying Cognitive Flexibility.” *ELife* 6 (March): e20552. https://doi.org/10.7554/eLife.20552.

Meder, David, Nils Kolling, Lennart Verhagen, Marco K. Wittmann, Jacqueline Scholl, Kristoffer H. Madsen, Oliver J. Hulme, Timothy E.J. Behrens, and Matthew F.S. Rushworth. 2017. “Simultaneous Representation of a Spectrum of Dynamically Changing Value Estimates during Decision Making.” *Nature Communications* 8 (1): 1942. https://doi.org/10.1038/s41467-017-02169-w.

Minoshima, Saroshi, Bruno Giordani, Stanley Berent, Kirk A. Frey, Norman L. Foster, and David E. Kuhl. 1997. “Metabolic Reduction in the Posterior Cingulate Cortex in Very Early Alzheimer’s Disease.” *Annals of Neurology* 42 (1): 85–94. https://doi.org/10.1002/ana.410420114.

Nassar, M. R., R. C. Wilson, B. Heasly, and J. I. Gold. 2010. “An Approximately Bayesian Delta-Rule Model Explains the Dynamics of Belief Updating in a Changing Environment.” *Journal of Neuroscience* 30 (37): 12366–78. https://doi.org/10.1523/JNEUROSCI.0822-10.2010.

Nestor, P. J., T. D. Fryer, M. Ikeda, and J. R. Hodges. 2003. “Retrosplenial Cortex (BA 29/30) Hypometabolism in Mild Cognitive Impairment (Prodromal Alzheimer’s Disease).” *European Journal of Neuroscience* 18 (9): 2663–67. https://doi.org/10.1046/j.1460-9568.2003.02999.x.

Pengas, George, John R. Hodges, Peter Watson, and Peter J. Nestor. 2010. “Focal Posterior Cingulate Atrophy in Incipient Alzheimer’s Disease.” *Neurobiology of Aging* 31 (1): 25–33. https://doi.org/10.1016/j.neurobiolaging.2008.03.014.

Spitmaan, Mehran, Hyojung Seo, Daeyeol Lee, and Alireza Soltani. 2020. “Multiple Timescales of Neural Dynamics and Integration of Task-Relevant Signals across Cortex.” *Proceedings of the National Academy of Sciences* 117 (36): 22522–31. https://doi.org/10.1073/pnas.2005993117.

Sun, Weilun, Ilseob Choi, Stoyan Stoyanov, Oleg Senkov, Evgeni Ponimaskin, York Winter, Janelle M. P. Pakan, and Alexander Dityatev. 2021. “Context Value Updating and Multidimensional Neuronal Encoding in the Retrosplenial Cortex.” *Nature Communications* 12 (1): 6045. https://doi.org/10.1038/s41467-021-26301-z.

Tan, Rachel H, Stephanie Wong, John R Hodges, Glenda M Halliday, and Michael Hornberger. 2013. “Retrosplenial Cortex (BA 29) Volumes in Behavioral Variant Frontotemporal Dementia and Alzheimer’s Disease.” *Dement Geriatr Cogn Disord*, 7.

Vann, Seralynne D., John P. Aggleton, and Eleanor A. Maguire. 2009. “What Does the Retrosplenial Cortex Do?” *Nature Reviews Neuroscience* 10 (11): 792–802. https://doi.org/10.1038/nrn2733.

Wilson, Robert C., Matthew R. Nassar, and Joshua I. Gold. 2013. “A Mixture of Delta-Rules Approximation to Bayesian Inference in Change-Point Problems.” Edited by Tim Behrens. *PLoS Computational Biology* 9 (7): e1003150. https://doi.org/10.1371/journal.pcbi.1003150.

Yu, Angela J., and Peter Dayan. 2005. “Uncertainty, Neuromodulation, and Attention.” *Neuron* 46 (4): 681–92. https://doi.org/10.1016/j.neuron.2005.04.026.