Okinawa Institute of Science and Technology
Graduate University

Thesis submitted for the degree

Doctor of Philosophy

The Role of Serotonin Neurons in Mouse Reward-based Behaviors

by

Masakazu Taira

Supervisor: **Kenji Doya**

January, 2022

Declaration of Original and Sole Authorship

I, Masakazu Taira, declare that this thesis entitled *The Role of Serotonin Neurons in Mouse Reward-based Behaviors* and the data presented in it are original and my own work.

I confirm that:

- No part of this work has previously been submitted for a degree at this or any other university.
- References to the work of others have been clearly acknowledged. Quotations from the work of others have been clearly indicated, and attributed to them.
- In cases where others have contributed to part of this work, such contribution has been clearly acknowledged and distinguished from my own work.
- None of this work has been previously published elsewhere.

Date: January, 2022
Signature:

Abstract

Serotonin (5-HT) is an important neuromodulator in reward-driven learning and decision making. The dorsal raphe nucleus (DRN) sends diffuse 5-HT projections throughout the brain. The involvement of DRN 5-HT neurons in reward-based behaviors has been examined using various types of behavioral tasks; however, how DRN 5-HT affects computational processes of decision making remains unclear. Reinforcement learning (RL) is a theoretical framework to describe the decision making process. Previous studies based on the RL framework have proposed hypotheses on the role of 5-HT in decision making, such as temporal discounting and model-based value computation. The overall aim of this thesis is to examine these hypotheses by analyzing behaviors under optogenetic manipulation, thereby clarifying the role of DRN 5-HT neurons in reward-based behaviors. The first hypothesis is that 5-HT modulates the relative importance of future rewards. Previous behavioral studies showed that 5-HT activation enhances patience to wait for future rewards and vice versa. However, how 5-HT regulates persistence to act for future rewards remains unknown. In the first part of my thesis research, I trained mice to perform a free-operant lever-pressing task, in which motor action rather than stationary waiting was required to obtain delayed rewards. In testing the effects of optogenetic activation and inhibition of 5-HT neurons on sustained motor actions, I found that optogenetic activation or inhibition of 5-HT neurons did not affect persistence in motor actions but an effect of the activation on slowing down response vigor, suggesting a different role of 5-HT neurons in motor actions for future rewards compared to stationary waiting. The second hypothesis examined is that 5-HT affects model-based decision making. In model-based decision making, agents use their own internal models of action-outcome relationships to plan forward and to select actions. Previous computational studies proposed facilitation of model-based decision making by 5-HT neurons, but behavioral evidence of how 5-HT regulates the process is still limited. A two-step decision making task is an established behavioral task to understand model-based decision making. In the second half of my thesis project, I trained mice to perform the two-step decision making task and found that optogenetic inhibition of 5-HT neurons affected choice behaviors and reduced time to make decisions possibly reflecting the disruption of model-based decision making. By fitting behavioral data to a model-free/model-based hybrid model, I found that photoinhibition of 5-HT neurons decreased the weight of model-based decision making. These results revealed the role of 5-HT neurons reward-based behaviors and model-based computations.

Acknowledgment

First of all, I would like to express my gratitude to my supervisor, Prof. Kenji Doya. When I joined OIST, I had no background in behavioral and computational neuroscience. Nevertheless, he patiently helped me to learn the field. Discussions with him always gave me ideas for experimental designs, data analysis, and interpretation. In addition, he provided me with many opportunities to talk with great scientists all over the world by inviting them to OIST or by letting me attend conferences. Without his supervision and kind support, this thesis would not have been possible. My deepest appreciations goes to my thesis committee members, Prof. Jeff Wickens and Prof. Bernd Kuhn. They read my proposal, research progress reports, and thesis, and gave helpful comments on them. Moreover, Prof. Wickens kindly allowed me to learn histology experiments in his lab for my thesis projects, and Prof. Kuhn provided helpful advise on academic career development.

I gratefully acknowledge support from lab members of the Neural Computation Unit. I am extremely grateful to Drs. Katsuhiko Miyazaki and Kayoko Miyazaki for supporting many aspects of my projects throughout my Ph.D. From Dr. Katsuhiko Miyazaki, I learned essential research techniques of animal surgery and behavioral experiments. Dr. Kayoko Miyazaki gave me much helpful feedback on my documents and management of my thesis projects. Thanks to their generous support and guidance, I remained motivated.

I also gratefully acknowledge help from Dr. Shiho Okitsu and Zhiguang Mu in training mice. Yuma Kajiwara offered helpful feedback on computational analyses of my data, and Mika Nishio and Tsukasa Miyake assisted with data collection and analysis. Christopher Buckley helped me assemble behavioral apparatus. I am grateful to research unit administrators, Mss. Emiko Asato, Kikuko Matsuo, and Misuzu Saito for arranging purchases of equipment and conference travel. Without the help from all the members of Neural Computation unit, I could not have completed my thesis project. I also would like to thank former lab members of the Neural Computation Unit, including Drs. Hiroaki Hamada, Tadashi Kozuno, Tomohiko Yoshizawa, Hiromichi Tsukada, and Carlos Enrique Gutierrez for interdisciplinary discussions about how to best spend my researcher life. Particularly, Drs. Hiroaki Hamada and Tadashi Kozuno supported and motivated me. Without their encouragement, I would not have survived my Ph.D. period.

I sincerely thank members of Neurobiology Research Unit led by Prof. Jeff Wickens, especially Dr. Julie Chouinard, Ms. Yumiko Akamine, and Mao-Ting Hsu for instructing me on basic skills necessary for histology and immunihistochemistry experiments. I also appreciate help from Dr. Yuichi Morohashi (OIST, Neuronal Mechanisms for

Critical Period Unit), Dr. Minzi Chan (Johns Hopkins University, US), and Ms. Ai Takahashi (OIST, imaging section) to improve my experimental protocols. For quantitative analysis of fluorescent image data, I express my appreciation for great help from Dr. Paolo Barzaghi (OIST, imaging section). I also would like to thank for Dr. Steven Aird for technical editing of my writings.

In addition, thanks go to an OIST graduate, Dr. Masakazu Igarashi. He gave me much feedback on my paper work. I also learned from him about how to develop research, and to decide future careers. Thank to his help, I have been able to carefully consider my future career as researcher.

I am deeply indebted to Dr. Thomas Akam, Marta Blanco-Pozo, and their supervisor, Prof. Mark Walton, at the University of Oxford. I had a great opportunity to visit their lab to learn the basic operations of the behavioral control system used for my project. They have also shared helpful information on experiment design and data analysis. Without their help, my second thesis project would not have been possible.

In addition, I thank my former supervisor, Prof. Hideki D. Kawai and the excellent Global Citizenship Program at Soka University. Without their support, I might have not done my Ph.D. at OIST. As a graduate of Soka University, I will keep in mind the counsel: 'For what purpose should one cultivate wisdom? May you always ask yourself this question!' and 'Only labor and devotion to one's mission give life its worth.'

Last, I thank my parents Kinue and Akikazu Taira for supporting me throughout my life. I think each of us HAS had difficulties during Ph.D. periods, some of which seemed too hard to overcome. However, my parents always encouraged me with their positivity. I am so proud to have overcome difficult times together and to be their son.

Abbreviations

ArchT	archaerhodopsin-T
ChR2	channelrhodopsin2
DA	dopamine
DRN	dorsal raphe nucleus
FR	fixed ratio
IPI	Inter-press interval
MAP	Maximum a posteriori
mPFC	medial prefrontal cortex
MRN	median raphe nuclues
NAc	nucleus accumbens
OFC	orbitofrontal cortex
PFC	prefrontal cortex
PR	progressive ratio
RL	reinforcement learning
SERT	serotonin transporter
SSRI	serotonin selective reuptake inhibitor
Tph	tryptophan hydroxylase
VTA	ventral tegmental area
5-HT	serotonin; 5-hydroxytryptamine

Contents

Declaration of Original and Sole Authorship iii

Abstract v

Acknowledgment vii

Abbreviations ix

Contents xi

List of Figures xiii

List of Tables xv

1 Introduction 1
- 1.1 Statement of the problem . . . 1
- 1.2 Outline of the thesis . . . 2

2 Review of literature 3
- 2.1 Biochemistry of 5-HT . . . 3
- 2.2 Neuroanatomy . . . 4
  - 2.2.1 Anatomical and chemical properties of DRN . . . 4
  - 2.2.2 Receptor distribution at DRN 5-HT projecting brain regions . . . 5
- 2.3 Reward inhibition and behavioral inhibition . . . 6
- 2.4 DRN 5-HT neural activity and reward/punishment-based learning and behavior . . . 7
- 2.5 Hypothetical roles of 5-HT neurons based on an RL framework . . . 7
  - 2.5.1 Hypothesis 1: Discount factor . . . 8
  - 2.5.2 Hypothesis 2: Model-based decision making . . . 12
- 2.6 Aims of my research . . . 15

3 Involvement of DRN 5-HT neurons in sustained motor actions for future rewards 17
- 3.1 Introduction . . . 17
- 3.2 Materials and Methods . . . 18
- 3.3 Results . . . 22

3.3.1 Photoactivation of DRN 5-HT neurons and behavioral tasks. . . 22
3.3.2 Activation of DRN 5-HT neurons prolonged stationary waiting for future rewards. 25
3.3.3 Activation of DRN 5-HT neurons neither enhanced nor suppressed persistence in motor actions. 26
3.3.4 Photoinhibition of DRN 5-HT neurons and behavioral tasks. . . 31
3.3.5 Inhibition of DRN 5-HT neurons shortened stationary waiting for future rewards. 33
3.3.6 Inhibition of DRN 5-HT neurons neither enhanced nor suppressed sustained motor actions 34
3.4 Discussion 38

4 Control of model-based decision making by DRN 5-HT neurons 41
4.1 Introduction 41
4.2 Materials and Methods 42
4.2.1 Behavioral experiment 42
4.2.2 RL model analysis 45
4.3 Results 49
4.3.1 Mice used both model-free and model-based decision making in the two-step task 49
4.3.2 Predicted effect of disrupted model-based decision making on choice behaviors 52
4.3.3 Photoinhibition and histological confirmation of optic probes . . 54
4.3.4 The effect of photoinhibition of DRN 5-HT neurons on first-step choice behaviors 54
4.3.5 The effect of DRN 5-HT photoinhibition on the decision-making process in the two-step task 56
4.3.6 The effect of photoinhibition of DRN 5-HT neurons on response time 61
4.4 Discussion 64

5 Conclusion and Future directions 67
5.1 Summary of findings 67
5.2 Limitations of the present study 68
5.3 Future directions 68

Bibliography 71

List of Figures

2.1 Biosynthesis pathway of 5-HT 4
2.2 Expression of 5-HT receptors in the prefrontal and mesolimbic pathways 6
2.3 Examples of two-step decision making tasks 15

3.1 Behavioral tasks and photoactivation protocol 23
3.2 The implantation site of optic probes in ChR2 mice. 24
3.3 DRN 5-HT photoactivation and stationary waiting for future rewards . 25
3.4 DRN 5-HT photoactivation and persistence in motor actions for future rewards 27
3.5 DRN 5-HT photoactivation and action vigor 30
3.6 Immunohistochemistry and the implantation site of optic probes in ArchT mice. 31
3.7 Behavioral tasks and photoinhibition protocol 32
3.8 DRN 5-HT photoinhibition and stationary waiting for future rewards . 33
3.9 DRN 5-HT photoinhibition and persistence in motor actions for future rewards 35
3.10 DRN 5-HT photoinhibition and action vigor 37

4.1 The schematic diagram of two-step decision making task 43
4.2 Behaviors in baseline sessions 51
4.3 Predicted effect of DRN 5-HT photoinhibition on first-step choice behaviors 53
4.4 The implantation site of optic probes in ArchT mice and the photoinhibition protocol 54
4.5 DRN 5-HT photoinhibition and first-step choice behaviors 55
4.6 Evaluation of RL models using the photoinhibition testing session data 56
4.7 DRN 5-HT photoinhibition and the weight of model-based decision making 58
4.8 DRN 5-HT photoinhibition and value learning/forgetting 60
4.9 Simulation of photoinhibition sessions using estimated parameter values 61
4.10 DRN 5-HT photoinhibition and response times 63

List of Tables

2.1 Previous studies on 5-HT regulation of motor actions 11
2.2 Previous studies on 5-HT control of model-based decision making . . . 15

4.1 Training schedule of each substage at training stage 4 44
4.2 Summary of the baseline session data. 49
4.3 MAP estimate values in the hybrid model in baseline sessions. 52
4.4 Summary of the photoinhibition testing session data. 54

Chapter 1

Introduction

We constantly make decisions. Decisions can be about simple issues such as which route to take on the way home, or whether to keep working to finish a task, or whether to take a break now. Sometimes we need to make critical decisions such as whether to quit a current job or to continue. In the process of decision making, we often select one among several options. To make a choice, we first estimate how much value each option could bring and choose the one with the greatest anticipated value. After the choice, we see how satisfying or disappointing the result was, and learn about the selected option. By repeating this process, we can make optimal decisions to maximize long-term returns. Disruption of this cognitive process causes inappropriate decision making and can produce clinical symptoms of psychiatric disorders [16, 98]. Understanding how the brain regulates decision making and the underlying computational process is critical not only to understand basic mechanisms of how the brain realizes adaptive behaviors, but also to understand the biological mechanisms underlying psychiatric disorders.

Serotonin (5-HT) serves multiple behavioral functions such as motor activity [32, 164], emotion [25], mood [118], motivation [103, 204], and decision making [83]. This molecule is also important as a clinical target. For example, selective serotonin reuptake inhibitors (SSRIs) are common drugs for treatment of affective disorders. Many previous studies have examined the role of 5-HT neurons in adaptive behaviors. However, it remains unclear how 5-HT affects reward-based learning and decision making.

1.1 Statement of the problem

The overall aim of this thesis is to better understand the role of 5-HT neurons in reward-based adaptive behavior. Reinforcement learning (RL) is a theoretical framework to describe how agents learn actions to maximize long-term rewards through experiences with actions and outcomes [51, 171]. Previous computational studies have hypothesized the role of 5-HT neurons in reward-based behaviors based on the RL framework, such as evaluating short- and long-term outcomes or balancing between reactive or deliberative decisions. However, some hypotheses are supported by limited behavioral evidence. Therefore, the specific aim of this thesis was to examine the role of 5-HT neurons in the dorsal raphe nucleus (DRN) in reward-based adaptive behaviors in mice and to biologically verify RL-based hypotheses.

1.2 Outline of the thesis

In this thesis, Chapter 2 reviews the role of 5-HT neurons in reward-based adaptive behaviors. Primarily, the role of 5-HT in punishment has been investigated. However, recent functional, pharmacological, and optogenetic studies have suggested that 5-HT neurons regulate reward-based learning and decision making, leading to hypotheses based on the RL framework. After reviewing general background information and previous behavioral studies of 5-HT systems, I introduce two hypotheses based on the RL framework. Finally, I raise additional questions and discuss the overall purpose of my experiments. In Chapter 3, I focus on the hypothesis that DRN 5-HT neurons control the relative importance of immediate versus future rewards. In particular, I describe the involvement of DRN 5-HT neurons in sustained motor actions for future rewards. In Chapter 4, I examine the hypothesis that DRN 5-HT neurons promote model-based decision making. In Chapter 5, I conclude the thesis by summarizing my findings and describing limitations and possible future work.

Chapter 2

Review of literature

2.1 Biochemistry of 5-HT

5-HT is synthesized from tryptophan through two chemical reactions [8]. First, tryptophan is converted into 5-OH tryptophan by tryptophan hydroxylase (Tph). By this reaction, a hydroxyl group is added to the benzene ring in tryptophan (From tryptophan to 5-hydroxytryptophan in Fig. 2.1). Next, the carboxyl group in 5-hydroxytryptophan is removed by tryptophan carboxylase (AADC in Fig. 2.1), and 5-HT is synthesized (From 5-Hydroxytryptophan to 5-HT in Fig. 2.1). The first chemical reaction from tryptophan to 5-hydroxytryptophan is the rate-limiting step. Previously, it was thought that 5-HT is produced outside the brain because Tph was only found in the peripheral nervous system. However, in 2003, it was shown that an isoform of Tph is also expressed in the brain [189]. After this finding, Tph was classified as Tph1 in the peripheral nervous system and Tph2 in the brain. 5-HT can be degraded by monoamine oxidase. Of two isoforms of monoamine oxidase, MAO-A and MAO-B, MAO-A has higher affinity for 5-HT than MAO-B [18]. 5-HT is mainly degraded into 5-hydroxyindole acetic acid by MAO-A and aldehyde dehydrogenase (from 5-HT to 5-hydroxyindole acetic acid in Fig. 2.1).

After synthesis, 5-HT is transported to axon terminals and packaged into synaptic vesicles via vesicular monoamine transporter 2 [160]. After the release of 5-HT-containing synaptic vesicles, 5-HT activates 5-HT receptors. 5-HT can act on 15 types of 5-HT receptors in the mammalian brain [14, 55, 84]. All 5-HT receptors except 5-HT3 receptors are G-protein coupled receptors. The 5-HT1 and 5-HT5 receptor families inhibit the activity of adenylyl cyclase (AC) via Gi/o and reduce the level of cyclic adenylyl monophosphate (cAMP). In contrast, 5-HT4, 5-HT6, and 5-HT7 receptors activate AC via Gs and increase the level of cAMP. 5-HT2 receptors activate phospholipase C via Gq and increase the level of diacylglycerol (DAG) and inositol 1,4,5 triphosphate (IP3), which leads to an increase in the intracellular Ca^{2+} concentration [105]. In addition to activation of 5-HT receptors, 5-HT released at the synaptic cleft is taken up by serotonin reuptake transporter (SERT), expressed in presynaptic 5-HT neurons [160].

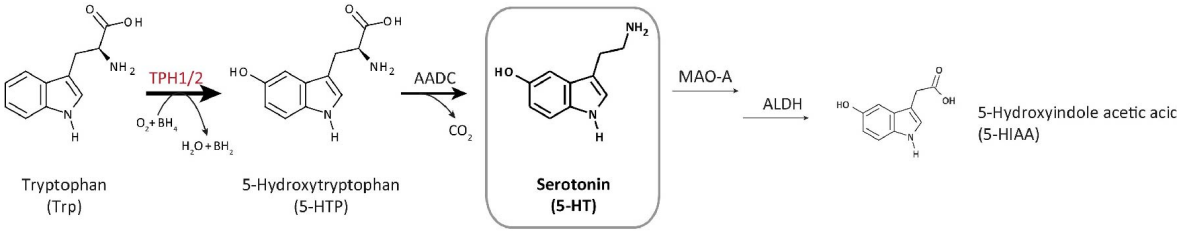

Figure 2.1: Biosynthesis pathway of 5-HT. The figure is adapted from [111] with modification.

2.2 Neuroanatomy

2.2.1 Anatomical and chemical properties of DRN

5-HT projections originate from the raphe nuclei located in the midbrain. In particular, the median raphe nucleus (MRN) and dorsal raphe nucleus (DRN) are the most rostral among the nine raphe nuclei. These two nuclei densely send diffuse projections to forebrain and midbrain regions [153]. Among them, the DRN has been examined as a central region sending 5-HT to the forebrain [65, 90]. The DRN has various neurotransmitters beside 5-HT. A subset of DRN 5-HT neurons co-releases different neurotransmitters, including glutamate [81, 102, 149, 153, 163, 191], GABA [166], and neuropeptides [153]. Several previous studies examined 5-HT-glutamate co-expressing neurons in the DRN [152, 191]. 60% of DRN 5-HT neurons co-release glutamate and express glutamate transporter type 3 [152]. Major interneurons in DRN are GABAergic, and they send local projections to 5-HT neurons to create a negative feedback loop [101, 168]. Some DRN neurons also express dopamine (DA) [100, 112].

DRN neurons receive dense inputs from cortical and subcortical regions and send outputs to these regions [103]. DRN 5-HT neurons receive inputs from the prefrontal cortex (PFC), including the insular, orbital, prelimbic/cingulate, and infralimbic cortices [7, 139, 165]. DRN 5-HT neurons also receive dense inputs from multiple subcortical regions, including the lateral habenula, amygdala, preoptic area, lateral hypothalamus, substantia nigra, and ventral tegmental area (VTA) [135, 139, 141, 182, 198]. The pattern of projection onto DRN 5-HT neurons depends on the originating regions. For example, the lateral habenula and PFC send glutamatergic inputs to DRN 5-HT neurons and control their activity via a feedforward inhibition circuit. In contrast, the lateral hypothalamus and amygdala send glutamatergic and GABAergic inputs and control 5-HT neurons with a push-pull mechanism [211]. DRN 5-HT neurons, in return, send wide projections to both cortical and subcortical regions. The PFC, including the cingulate, prelimbic, infralimbic, and orbital cortices receives projections [67, 139]. DRN 5-HT neurons also send projections to the mesocortical or dopaminergic pathway, including the VTA [78, 181], striatum [52, 202], and other subcortical regions such as the amygdala and hypothalamus [152, 169]. In relation to chemical characteristics, 5-HT-glutamatergic neurons are located in the dorsomedial part of DRN and mainly project to cortical regions [152]. Some DRN 5-HT neurons corelease glutamate in the VTA [191].

Recent virus tracing studies revealed detailed anatomical characteristics of DRN 5-HT neurons [152, 153]. They first examined the topographical distribution of 5-HT neurons projecting to different brain areas. 5-HT neurons send dense projections to subcortical areas from the dorsal part of the DRN and other 5-HT neurons project to cortical areas mainly from the ventral part [152]. This study also found that axonal collateralization of cortical and subcortical projections are segregated and inputs to these two DRN 5-HT neurons are biased. Also, single-cell-level projection reconstruction suggests that DRN 5-HT neurons have several different patterns of axon collateralization [153]. These anatomical studies showed that DRN 5-HT projections are anatomically segregated and diverse.

2.2.2 Receptor distribution at DRN 5-HT projecting brain regions

Various types of 5-HT receptors are distributed across different brain regions and different compartments of a single neuron, which enables 5-HT receptors to work in functionally different ways. In particular, the roles of 5-HT1A, 1B, 2A, 2B, and 5-HT3 receptors in adaptive behaviors have been extensively examined [76]. In PFC, 5-HT1A and 5-HT2A receptors are mainly expressed in pyramidal neurons [143, 144, 158, 196]. 5-HT1A receptors inhibit activity in the axon initial segment via suppression of action potential generation [37, 39, 70, 147], while 5-HT2A receptors excite neuronal activity at apical dendrites [91, 108, 109]. 5-HT2C receptors can also be found in PFC pyramidal neurons [148]. PFC interneurons also express multiple types of 5-HT receptors. 5-HT1A and 5-HT2A receptors are expressed in fast-spiking interneurons in the deeper layers of the PFC [148], and both receptors are expressed separately in different neurons [129]. On the other hand, interneurons that express 5-HT3 receptors are located in superficial layers of the PFC [53, 188]. These interneurons regulate the function of PFC pyramidal neurons. 5-HT also acts on 5-HT receptors in dopaminergic pathways, which results in regulation of DA release [4]. In the striatum, 5-HT1B, 5-HT2A, and 5-HT2C receptors are expressed abundantly [52, 202]. Thus, in the nucleus accumbens (NAc), DA release is facilitated by 5-HT3 receptors [26, 45], while in the ventral and dorsal striatum, DA release is inhibited by 5-HT2C receptors [5]. 5-HT1B receptors are expressed in medium spiny neurons in the striatum, and control lateral inhibition between medium spiny neurons [142]. 5-HT receptors are also expressed in VTA neurons. VTA DA neurons express 5-HT2A, 5-HT2C, and 5-HT3 receptors. Within the VTA, GABA neurons express 5-HT2A and 2C receptors. Previous studies showed that 5-HT2C receptors in the VTA regulate DA neural activities and DA release in the striatum[23, 24]. DA release is inhibited by 5-HT2C receptor agonists possibly via a disrupted balance between VTA DA and GABA neural activities [56, 133]. On the other hand, the 5-HT3 receptor in VTA mediates DA release at NAc [191]. Apart from the above-mentioned expression of 5-HT receptors in DRN projecting regions, 5-HT1A receptors are also expressed in the soma and dendrites of DRN 5-HT neurons as autoreceptors to suppress 5-HT neuron activity [127, 184]. Expression of 5-HT receptors in the prefrontal and mesolimbic pathways is summarized in Figure 2.2.

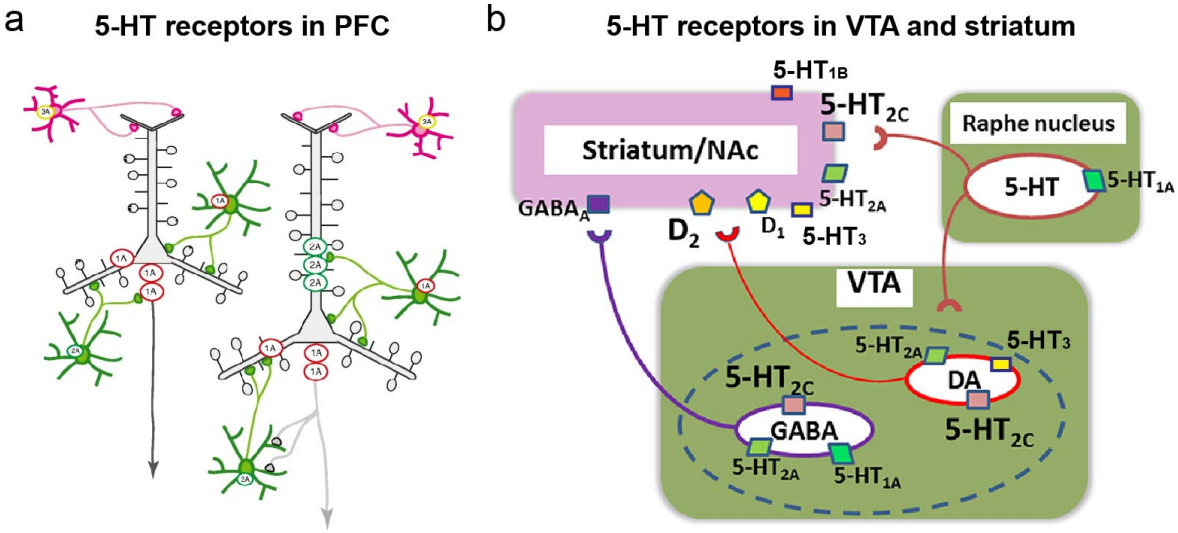

Figure 2.2: Expression of 5-HT receptors in prefrontal and the mesolimbic pathways a: 5-HT receptor expression in the PFC (adapted from [146]) b: 5-HT receptor expression in the striatum and VTA (adapted from [170] with modification).

2.3 Reward inhibition and behavioral inhibition

Based on behavioral experiments combined with pharmacology, genetics, and electrophysiology, several hypotheses have been proposed to clarify the role of 5-HT neurons in adaptive behaviors. The first theory is that 5-HT regulates responses to punishment or aversive information by antagonizing DA neurons, the so-called 'punishment' or 'reward inhibition' theory [40, 44, 46, 167]. In terms of the response to punishment, previous studies showed that aversive electric shock increased 5-HT neuron activity [6, 172], and depletion of 5-HT neurons reduced the response to foot shock [179]. Some other studies have shown that increased 5-HT levels reduced anxiety [72, 73]. 'Reward inhibition' theory is also supported by studies showing that 5-HT neurons inhibit the response to reward stimuli using intracranial self-stimulation and a conditioned place preference test [57–59, 177].

The second theory is the 'behavioral inhibition' theory [19, 30, 34, 43, 44, 167]. This theory hypothesizes that increased 5-HT activity induces suppression or inhibition of behaviors that lead to punishment. In human studies, depletion of dietary tryptophan (the precursor of 5-HT) to lower 5-HT levels disrupted punishment-induced response inhibition without affecting general motor activity [35, 36]. A recent human study measured 5-HT signals from human patients with Parkinson's disease using fast-scan cyclic voltammetry and found that 5-HT signals negative prediction error. Positive signals result when subjects get smaller rewards than expected, and its fluctuations in 5-HT were positively correlated with protective action from loss [130]. These studies support the behavioral inhibition hypothesis in the context of punishment.

2.4 DRN 5-HT neural activity and reward/punishment-based learning and behavior

The theories discussed above focus on the role of 5-HT neurons in the context of aversiveness or inhibition of reward effects. However, recent electrophysiological and imaging studies indicate that DRN neural activity changes in response to multiple variables. Electrophysiological studies in primates showed that DRN neural activity was modulated by expected and received reward size [21, 86, 132]. Another primate electrophysiology study showed that DRN neurons respond differently to sound cues that signal different types of expected appetitive or aversive outcomes and their probabilities [75]. In another study using rat electrophysiological recordings during a two-odor discriminative task, DRN neurons fired in a phasic manner to sensory cues about expected rewards, selection of actions leading to rewards, and reward acquisition [150]. Those previous studies demonstrated coding of diverse information by DRN neurons, but did not identify 5-HT neurons. Recent studies selectively recorded DRN 5-HT neural activity by labeling 5-HT neurons with genetically encoded calcium indicators or optogenetically identifying DRN 5-HT neurons [29, 99, 110, 209]. Pavlovian conditioning tasks have been used to examine how 5-HT neurons are modulated by expected reward/punishment [29, 110, 209]. Unit recording in a mouse Pavlovian conditioning task indicates that 5-HT neurons encode predictive reward/punishment with both phasic and tonic excitation at multiple timescales [29]. 5-HT neural activity was modulated by unexpected events related to the uncertainty of future reward/punishment [110]. Also, DRN 5-HT neural activities changed dynamically while mice were forming predictive responses to the unconditioned stimulus [209]. At first, phasic firing of 5-HT neurons was found upon reward acquisition, but tonic modulation was formed between predictive cues and rewards. Several studies have conducted electrophysiology or fiber photometry in DRN neurons during foraging tasks with delayed rewards [99, 122]. The fiber photometry study showed that tonic 5-HT neural activity was enhanced during the reward expectation period and phasic excitation was observed at reward acquisition [99]. These studies suggest that DRN 5-HT neurons are dynamically modulated during reward-driven behaviors. Tonic activation of putative DRN 5-HT neurons during the reward expectation period was observed in another behavioral task with delayed rewards [122].

2.5 Hypothetical roles of 5-HT neurons based on an RL framework

The functional studies mentioned above raise the question of how 5-HT neurons regulate cognitive computation and reward-driven adaptive behaviors. The RL framework proposed several hypothetical roles of 5-HT neurons [40, 44, 50, 124]. RL is a theoretical framework to describe the decision making process in reward/punishment-driven behaviors[51, 171]. RL agents try to maximize long-term rewards by learning optimal policies. To learn optimal policies, agents evaluate candidate actions in three steps. They estimate values of possible actions in a given state. They select an action based

on estimated values, and they learn the value of the state and action, given an observed outcome. How RL agents learn action values and guide actions is a critical issue. In my thesis, two hypothetical roles of 5-HT neurons in value learning and action selection are described in the following sections.

2.5.1 Hypothesis 1: Discount factor

Theoretical account

In order to maximize long-term rewards, an RL agent needs to take into account not only immediately accessible rewards, but also rewards expected in the distant future. In the RL framework, the discount factor controls the relative importance of future rewards in the current state. If the discount factor is smaller, the agent puts more weight on immediate rewards than delayed rewards. Therefore, an agent with a small discount factor is likely to prefer small immediate rewards over large delayed rewards. Previous studies have proposed that 5-HT controls the discount factor [50, 161]. According to this hypothesis, when 5-HT levels are higher, organisms are more likely to act so as to obtain delayed rewards.

Behavioral studies: Patience to wait for future rewards

Some previous behavioral and functional studies show consistent results with regard to this hypothesis. Previous studies using lesioning and pharmacological manipulation indicate that depletion of 5-HT neurons increased impulsiveness, which is characterized by increased premature responses to gain immediate rewards rather than delayed rewards [71, 123, 200]. Another study indicated that putative DRN 5-HT neural activity increased during waiting for future rewards in rodents [122]. In order to examine the behavioral role of DRN 5-HT neurons, the effect of downregulation of DRN 5-HT neurons was examined [127]. In this study, DRN 5-HT neural activities were suppressed by local infusion of a 5-HT1A agonist, which induced activation of 5-HT1A autoreceptors in DRN 5-HT neurons. Chemical downregulation of DRN 5-HT neurons increased premature responses to delayed rewards. In order to further causally examine the role of DRN 5-HT neurons in waiting for delayed rewards, DRN 5-HT neural activities were optogenetically activated while mice were performing a delayed reward task [128]. In this task, mice were required to keep poking their noses into a hole to obtain food rewards. During the task, reward omission trials in which the mice could not get food were probabilistically inserted. Duration of nose pokes in omission trials measured how much delay the mice could tolerate under the expectation of future rewards. DRN 5-HT neurons were activated during the waiting. I found that photoactivation of DRN 5-HT neurons increased waiting duration in reward omission trials [128]. Another optogenetic study also showed consistent results in waiting for delayed tones associated with rewards [63]. In a previous study using an appetitive Pavlovian conditioning task, optogenetic stimulation of the DRN modulated coding of future rewards in the orbitofrontal cortex (OFC) of mice [210]. In an intertemporal choice task, DRN 5-HT neurons bi-directionally controlled impulsive behaviors [136, 203]. A human fMRI study showed that prediction of larger rewards following immediate loss increased DRN neural activity [173]. When the 5-HT level was decreased by dietary regulation

of tryptophan in humans, the timescale for predicting future rewards was shortened and subjects preferred to choose small immediate rewards over large future rewards [162, 174]. These rodent and human studies support the discount factor hypothesis and also suggest that 5-HT neurons regulate patience to wait for future rewards.

Behavioral studies: Patience to act for future rewards

Previous studies investigated the discount factor hypothesis using a task in which subjects were required to wait for delayed rewards. However, in real life, we also encounter situations in which waiting for future rewards involves physical activity (e.g. keep climbing steep roads of Mt. Fuji to enjoy the views from the summit). Therefore, if 5-HT controls the discount factor, it is possible that 5-HT neurons regulate persistence to act for future rewards. Previous studies have examined the role of 5-HT in motor actions using pharmacological manipulation and behavioral tasks in which subjects are required to invest physical effort (e.g. pressing a lever, actively performing nose poke, and squeezing a hand grip). Previous human studies showed that chronic administration of SSRIs, which increased extracellular 5-HT concentrations, increased the monetary reward by reducing effort cost [117] and by enhancing learning of reward and effort dynamics [159].

In rodents, various experimental paradigms have been used. One of the behavioral tasks used in rodent studies is T-maze barrier climbing. In this task, rewards of different sizes are located in the two opposite arms. In the arm with the larger reward, barriers are located, but not in the arm with smaller rewards. The larger reward can be obtained by climbing the barrier located before the reward (large reward with high cost), while the smaller reward can be obtained without any effort or with less effort than the opposite arm (small reward with low cost). In previous studies in which rats performed this task, 5-HT depletion with a 5-HT synthesis blocker did not change the probability of selecting the large reward with high cost over the small reward with low cost [47, 89]. However, another study using the same task showed that the amount of reduction in SERT in the cingulate, somatosensory, and insular cortices induced by methamphetamine correlated with the probability of selecting a large reward with high cost, indicating the possible involvement of 5-HT neurons in motor actions [96].A concurrent food choice task is another type of task choosing between large rewards with high effort (highly preferred food gained by lever pressing) and small reward with low effort (freely available lab chow). Acute SSRI treatment reduced the number of lever-presses without changing the consumed amount of lab chow [205, 206]. Genetic deletion of SERT to increase 5-HT levels also reduced the lever-pressing response to obtain preferred food [157]. Running wheel activity has also been used to examine the effect of SSRIs on rewarding motor actions, where acute SSRI treatments decreased running in mice without affecting general locomotor activity [197].

In addition to the choice tasks mentioned above, the progressive ratio (PR) lever press task has been also used to examine motivated behaviors. In a PR task, the required number of lever presses to gain rewards increases as a session progresses. To quantify the incentives of reward and action persistence, an index called breakpoint (the number of lever presses completed before abandoning the session) is measured [154]. Administration of tryptophan to increase 5-HT levels did not change the breakpoint

[115], but genetic deletion of SERT or chronic treatment with SSRI reduced it [157]. On the other hand, previous studies using PR tasks consistently showed that 5-HT2C receptors regulate motor actions motivated for food rewards. Systemic injection of selective agonists of 5-HT2C receptors such as mCPP, lorcaserin, Ro 60-0175, and CP-809101 reduced the breakpoint [15, 62, 79, 80, 180, 193], and this effect was removed by co-administration of the 5HT2C receptor antagonist, SB-2420284 [79, 80]. Other studies also showed that the breakpoint was reduced by local infusion of a 5-HT2C receptor agonist to the VTA [61] but not to the NAc [145]. A recent chemogenetic study strongly supported these pharmacological studies, showing that selective activation of 5-HT2C receptors in the VTA reduced the breakpoint [180]. Another study showed that systemic injection of a 5-HT2C receptor antagonist alone increased the breakpoint [10]. These studies suggest that the 5-HT2C receptor modulates DA neurons and motor actions [11]. However, the use of breakpoints needs to be discussed further, because breakpoints are affected not only by incentives for rewards, but also by the required response ratio and step size of the ratio [95]. In order to assess only incentives for rewards and action persistence, a mathematical model [20] was applied to some previous studies. The mathematical principle of reinforcement (MPR) is the basis of this model [94]. By this mathematical model, it was shown that systemic injection of a 5-HT2C agonist reduced the breakpoint by impacting motor parameters, but not motivational parameters [15], while specific activation of VTA 5-HT2C receptors changed motivational parameters, resulting in reduction of the breakpoint [180]. This mathematical model was also used to examine the effect of the 5-HT1A agonist, 8-OHDPAT, and antagonist, WAY-100635, on the PR task. Here, a 5HT1A agonist increased incentives for rewards, but also impaired motor performance by affecting postsynaptic 5HT1A receptors, and that effect was removed by a 5-HT1A receptor antagonist [82, 208]. On the other hand, selective manipulation of 5-HT1B receptors did not change the breakpoint [60, 155]. These studies indicate that regulation of motor actions differs with subtypes of 5-HT receptors and how they manipulate 5-HT activity.

Although previous studies suggest the possible downstream effect of 5-HT release, they did not directly examine how DRN 5-HT neural activities are related to motor actions. In addition to behavioral studies, several electrophysiological studies have tried to examine the role of DRN 5-HT neurons in motor control. One previous study performed single-unit recording from putative DRN 5-HT neurons during treadmill-induced locomotion in cats. In this study, an increase in treadmill speed did not change the firing of DRN 5-HT neurons [183]. However, another study conversely found that subtypes of DRN 5-HT neurons are responsive to treadmill-induced locomotion [195]. In a task in which cats have to press a pedal to gain a reward, a large number of DRN neurons were either inhibited or excited before pressing a pedal, suggesting that 5-HT neural activity may be involved in motor actions for rewards [97]. A recent fiber photometry study monitored population activity from DRN 5-HT neurons. This study showed that DRN 5-HT group neural activities did not change during lever-pressing for rewards, compared to baseline neural activity [207]. On the other hand, it was also reported that under threatening conditions, such as under tail suspension or a forced swim test, DRN 5-HT neural activities are positively correlated with active behaviors [164].

A recent optogenetic study examined the role of DRN 5-HT neurons in behavioral activation. DRN 5-HT neurons were optogenetically activated while mice were performing a dynamic foraging task [104]. In this task, two foraging sites were located on opposite walls of an operant chamber and the mouse needed to poke its nose into the foraging sites to gain a water reward and to visit the two sites. A nose poke to a foraging site stochastically delivered a water reward, but repeated nose pokes exponentially decreased the probability of a reward. In a probabilistic foraging task, optogenetic activation of 5-HT neurons increased the number of nose pokes and lengthened the duration of stays in a foraging site, suggesting that activation of 5-HT neurons promotes action persistence [104]. In summary, previous studies using pharmacological and optogenetic manipulation showed the involvement of 5-HT in sustained motor actions (Table 2.1). However, it remains possible that the main observed effect is due to an interaction with DA neurons [22, 191].

Table 2.1: Previous studies on 5-HT regulation of motor actions

Task	Species	Manipulation	Effect	Reference
T-maze barrier climbing	Rats	5-HT synthesis inhibition	No change	[47, 89]
	Rats	Metaamphetamine	improved behavioral flexibility	[96]
Concurrent food choice	Rats	Acute SSRI	decreased lever-press	[205, 206]
	Mice	Chronic SSRI	decreased lever-press	[157]
	Mice	SERT knockout	decreased lever-press	[157]
Running wheel	Mice	Acute SSRI	Increased response	[197]
PR lever-press	Rats	Systemic tryptophan	No change	[115]
	Mice	Chronic SSRI	Reduced breakpoint	[157]
	Mice	SERT knockout	Reduced breakpoint	[157]
	Rats	Systemic 5-HT2CR agonist	Reduced breakpoint	[15, 62, 79, 80, 180]
	Mice	Systemic 5-HT2CR agonist	Reduced breakpoint	[180, 193]
	Mice	Chemogenetic VTA 5-HT2CR activation	Reduced breakpoint	[180]
	Rats	5-HT2CR agonist in VTA	Reduced breakpoint	[61]
	Rats	5-HT2CR agonist in NAc	Reduced breakpoint	[145]
	Mice	Systemic 5-HT2CR antagonist	Increased breakpoint	[10, 11]
	Rats	Systemic 5-HT1AR agonist/antagonist	Increased/reduced breakpoint	[82, 208]
	Rats	5-HT1BR agonist in NAc	No change	[60]
	Mice	5HT1BR knockout	No change	[155]
FR lever-press	Mice	DRN 5-HT photoactivation + systemic SSRI	increased lever-press	[22]
Probabilistic foraging	Mice	DRN 5-HT photoactivation	increased active nose-poke	[104]

2.5.2 Hypothesis 2: Model-based decision making

Theoretical account

There are several ways to learn action values and to choose an action. In model-based decision making, the agent has its own internal model regarding the action-outcome relationship. In *shogi*, Japanese chess, the internal model corresponds to imagining "How will an opponent move if I take this action?" Using the internal model, the agent mentally simulates what is likely to happen next by taking actions and selects actions based on the mental simulation. In some studies, model-based decision making is also referred to as planning, because the agent plans forward using its internal model to select actions. On the other hand, in model-free decision making, the agent does not use an internal model, but instead relies entirely on past experience with actions and observed outcomes. In the example of *shogi*, if the agent guides its actions using model-free decision making, the agent chooses an action based on whether the action worked previously. Previous computational accounts proposed that 5-HT neurons are involved in model-based value learning and decision making [43, 124]

Behavioral studies: 5-HT and model-based decision making

Model-based decision making has been examined in several experimental paradigms. The first experimental paradigm is with a serial reversal contingency. In this paradigm, after forming a predictive association between action and reward, this contingency was switched without explicit instruction. Using this type of task, previous studies examined the role of 5-HT neurons in rats [12, 13], primates [27, 28], and humans [93]. These studies consistently showed that depletion of 5-HT disrupted behavioral adaptation to new contingency. However, these studies did not produce direct evidence that the underlying computation is really accompanied by model-based decision making, value learning based on the internal model of action-outcome contingency. Rather, a recent study showed that mice choose behaviors in a probabilistic task with a reversal of action-reward contingency that can be captured in a model-free RL framework [85]. Alternative paradigms can more directly assess model-based decision making.

One of the alternative methods is the outcome devaluation task. In the first instrumental conditioning stage, animals or human subjects learn contingencies between actions and outcomes. For example, in rodent studies, they learn to press a lever or poke their noses into a hole in order to obtain food rewards. After the initial training, rewards associated with learned actions are devalued by paring them with sickness or inducing satiation. In the test phase, animals are placed in the behavioral box. If animals respond based on a model-based system, they infer that having more responses will lead to devalued food. Through such an inferential process using the model of action and consequences, model-based behavioral control prevents animals from responding by pressing a lever or nose poking. One rat study showed that the instrumental response was decreased in wild-type rats, but not in SERT knockout rats [134]. This suggested involvement of the 5-HT system in model-based decision making. Also, a recent mouse study showed a causal relationship between DRN 5-HT neurons and model-based decision making [137]. In this study, mice were trained to perform nose pokes to obtain food rewards and food was devalued by pairing it with

lithium-induced sickness. In the test phase, DRN and MRN 5-HT neural activities were optogenetically silenced. Optogenetic silencing of DRN 5-HT neurons increased nose poke responses compared to controls after reward devaluation. On the other hand, inhibition of MRN 5-HT neurons did not change the response. This study suggests that DRN 5-HT neurons regulate prospective inference using the internal model of the action-outcome consequence and decision making guided by the model-based inference.

Interestingly, recent behavioral and computational studies on waiting behavior also suggest that DRN 5-HT neurons affect model-based decision making. As mentioned above, the role of DRN 5-HT neurons in patience to wait has been examined with the hypothesis that DRN 5-HT neurons control the discount factor in model-free decision making. However, new findings on DRN 5-HT regulation of waiting behaviors are difficult to interpret using the former hypothesis. A recent study reported that the effect of DRN 5-HT activation was different by the uncertainty of reward timing and probability [124]. Photoactivation of DRN 5-HT neurons had a larger effect on waiting duration in omission trials when the delays of rewarding trials were variable than when they were fixed. Also, The effect of photoactivation was larger when the reward probability (i.e. probability of rewarding trials) was high, e.g., 75% rewarding trials and 25% omission trials, than when it was low, e.g., 25% rewarding trials and 75% omission trials. To describe this effect, the study proposed a Bayesian decision model of waiting. In this model, it was assumed that mice have their own internal model about the probability distribution of reward timing and also infer hidden states of whether the current trial is a reward or an omission trial. While waiting, mice update their belief about the current trials based on the observation that they cannot obtain rewards at the present timing. The posterior belief was updated based on a prior belief about the current trial and the likelihood of a reward trial at the present time, which was calculated using the internal model of reward timing. In other words, model-based value learning occurred to update the belief about the current trial. The belief represents the action value "keep waiting." Based on this computational model, it was proposed that DRN 5-HT activation increased prior belief of a reward trial, resulting in waiting. This study also suggested that DRN 5-HT neurons encode latent decision variables, the subjective belief of reward in this study, to affect model-based value learning.

In actual action selection, both model-free and model-based decision making systems work in parallel [41, 48, 49]. This is also evident in the reward devaluation task mentioned in a previous paragraph. If mice simply rely on model-based decision making, it is expected that no response would be taken. However, mice respond to some extent even after reward devaluation. This suggests that mice tend to use both values learned by model-based and model-free decision making systems. In order to examine how these two valuation systems work together to make a decision, a two-step decision-making task was developed for experiments with human subjects ([42]; Fig. 2.3a). This task consists of two decision-making points. After the first-step choice, the state moves to either of two second-step states probabilistically. The state transition probability is non-neutral. In other words, one of the first-step actions commonly leads to one second-step state and rarely to the other, and vice versa (See the left drawing of Fig. 2.3a). In each state, two choices were presented. By making a second-step choice, a reward was given probabilistically. That simulation study suggested that the choice

behavior of this task can be differentiated between model-free RL and model-based RL agents ([2, 42]; See two bar graphs on the right side of Fig. 2.3a). Model-free RL agents reinforce their first-step actions taken after rewarded trials and are likely to repeat the same first-step choice after a reward regardless of the trial. On the other hand, model-based RL agents estimate values of first-step actions based on transition probability from first-step choices to second-step states. Therefore, when agents are rewarded after rate transition, they behave as though "mentally simulating" each possible state action pair and trying to choose the action which commonly leads to the second-step state where they were rewarded in the preceding trial. As a result, model-based RL agents are likely to switch first-step choices after a trial with a reward and a rare transition. To summarize, in the two-step decision making task, the model-free decision making system reinforces actions directly from experiences of rewards, while the model-based decision making system leads to selecting actions by values based on outcome and transition probability, and they show different patterns of choice behaviors.

A previous human study used this behavioral task to examine whether the 5-HT system controls arbitration between model-free and model-based RL systems [201]. They reduced the 5-HT level of healthy human subjects by acute tryptophan depletion. The study showed that acute tryptophan depletion increased the weight of the model-free RL system to guide actions. They also recently determined the correlation between the weight of RL systems and the amount of SERT in brain regions. In detail, the weight of the model-based RL system is positively correlated with the amount of SERT in OFC. On the other hand, the weight of the model-free RL system is negatively correlated with that in the putamen, suggesting a detailed neural substrate possibly affected by 5-HT systems in arbitration of the two systems [187]. The advantage of this task is that it is possible to clearly demonstrate which computation for decision making is disrupted by intervention by analyzing choice behaviors with RL models. More importantly, several previous studies developed similar two-step tasks for rats [68, 74, 119] and mice ([3]; Fig. 2.3b). Most previous rodent studies suggest that animals also use both model-free and model-based systems to make the first-step choice action. Particularly in the case of the mouse study, the authors monitored neural activities in the anterior cingulate cortex and tested the effect of optogenetic inhibition, suggesting that use of the rodent two-step task will lead us to understand neural correlates with latent decision-making variables, as well as the causal relationship between decision making and neural activities.

In summary, previous behavioral and computational studies suggest involvement of the 5-HT system in model-based decision making (Table 2.2). Starting from systemic manipulation of 5-HT levels, the rodent study specifically identified the importance of DRN 5-HT neurons, rather than MRN 5-HT neurons. However, it remains unclear how DRN 5-HT neurons modulate computational processes of model-based decision making. As mentioned above, by combining a two-step decision-making task and optogenetic manipulation, it is possible to examine regulation of specific cell-types in computational processes of model-based decision making.

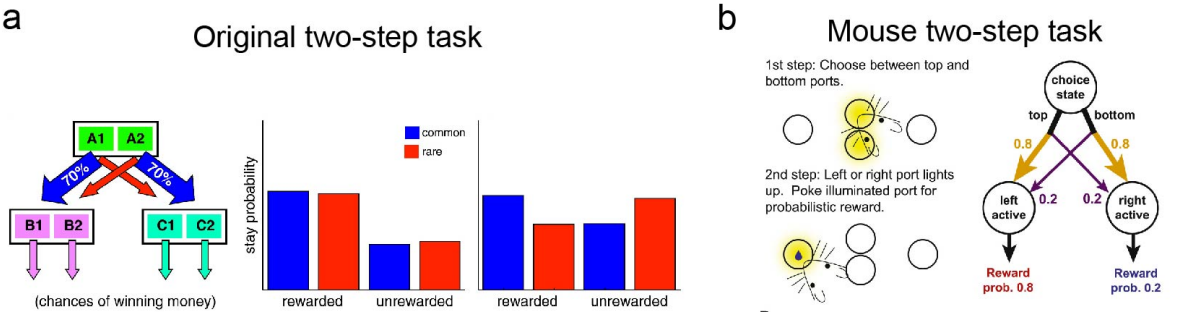

Figure 2.3: Two-step decision making task a: A human two-step decision making task adapted from [42]. The left figure is a schematic drawing of a trial. The middle and right graphs indicate stay probabilities simulated by model-free RL and model-based RL agents, respectively. b: A mouse two-step decision making task adapted from [3]. The left figure is a schematic drawing of a trial. The right diagram indicates a state transition in a trial.

Table 2.2: Previous studies on 5-HT control of model-based decision making

Task	Species	Manipulation	Effect	Reference
Reversal learning	Rats	forebrain 5-HT depletion	decreased behavioral flexibility	[12]
	Rats	systemic SSRI	improved behavioral flexibility	[13]
	Marmosets	Prefrontal 5-HT depletion	decreased behavioral flexibility	[27, 28]
	Human	Acute tryptophan depletion	decreased behavioral flexibility	[93]
	Rats	5-HTT knockout	Increased response	[134]
	Mice	DRN 5-HT photoinhibition	Increased response	[137]
	Mice	MRN 5-HT photoinhibition	Increased response	[137]
Two-step task	Human	Acute tryptophan depletion	shift toward model-free behavior	[201]

2.6 Aims of my research

Based on these previous studies, the overall motivation of my thesis was to better understand the role of DRN 5-HT neurons in reward-driven behaviors and underlying computational processes. In Chapter 3, the first hypothesis, that DRN 5-HT neurons control the relative importance of future rewards, was examined in a behavioral task requiring sustained motor actions for future rewards. The effect of optogenetic manipulation of DRN 5-HT neurons was tested. Next, in Chapter 4, I examine the other hypothesis that DRN 5-HT neurons promote model-based decision making. For this purpose, mice were trained to perform a two-step decision-making task and I tested the effect of optogenetic inhibition of DRN 5-HT neurons on choice behavior and underlying computational processes.

Chapter 3

Involvement of DRN 5-HT neurons in sustained motor actions for future rewards

3.1 Introduction

Previous computational studies based on the RL framework proposed that 5-HT controls the temporal discount factor and that activation of 5-HT neurons increases the relative importance of future rewards over immediate rewards [50, 161]. In support of this hypothesis, a series of experimental studies using delayed reward tasks has shown increased 5-HT transmission while rats were waiting for delayed rewards [122, 126]. Furthermore, pharmacological inhibition of DRN 5-HT neurons increased premature abandonment of delayed rewards[127] and optogenetic activation of those neurons prolonged the time spent for waiting for delayed rewards, establishing a causal relationship between DRN 5-HT neurons and patience in waiting for future rewards [63, 124, 125, 128]. While these studies examined the role of DRN 5-HT neurons in passive waiting to obtain future rewards, how they regulate active behavior to obtain future rewards has not been well studied, except in the context of patch-leaving decision making [104].

To examine the role of DRN 5-HT neurons in sustained motor actions, I trained mice to perform an operant conditioning task that requires variable numbers of lever-presses. Then, I tested the effect of optogenetic activation and inhibition of DRN 5-HT neurons on motor actions. For comparison, I also trained the same mice for a stationary waiting task and tested the effect of optogenetic manipulation of DRN 5-HT neurons. I found that optogenetic activation of DRN 5-HT neurons prolonged stationary waiting for future rewards, whereas optogenetic inhibition reduced the waiting time. On the other hand, optogenetic activation or inhibition of DRN 5-HT neurons had no effect on persistence of motor actions, suggesting that DRN 5-HT neurons regulate two types of behaviors for future rewards in different ways.

3.2 Materials and Methods

Animals. All experimental procedures were performed in accordance with guidelines established by the Okinawa Institute of Science and Technology Experimental Animal Committee. For optogenetic activation experiment, I used eight Tph2-ChR2(C128S)-EYFP bi-transgenic mice. ChR2(C128S) is a step-type opsin that remains activated by a short pulse of blue light and deactivated by yellow light [128]. Among the eight Tph2-ChR2 mice, four mice were first tested in the lever-pressing task and then the stationary waiting task, and the other four mice were tested in opposite order. For controls, five Tph2-tTA transgenic mice were used. All Tph2-tTA mice were first tested in the lever-pressing task and then the waiting task.

For the optogenetic inhibition experiment, I used four Tph2-ArchT-EGFP bi-transgenic mice. ArchT activates an inhibitory current in response to yellow light. Blue light was used for controls, as in previous studies [178, 207]. All Tph2-ArchT mice were first tested in the waiting task and then in the lever-pressing task.

All mice were housed individually at 24°C on a 12:12 h light: dark cycle (lights on 07;00-19;00 h). All behavioral training and testing sessions were performed during the light cycle, 5 days per week. Mice were deprived of access to food one day prior to the first training session and could acquire food during training and testing sessions. Food was freely available during days off until 24 h before the next session. Mice could freely access water in their home cages.

Behavioral apparatus. All training and testing sessions were performed in operant boxes (Med-associates, 21.6 cm width x 17.8 cm depth x 12.7 cm height). Two 2.5 cm square holes were located in the walls on opposite sides of the box. One hole was designated as the reward site connected to a food dispenser delivering 20-mg food pellets, while the other hole was defined as a tone site. A retractable lever was positioned to the left of the reward site. One 2.8-W house light and one speaker were located above and to the upper right of the tone site, respectively. Hardware attached to the operant boxes was controlled via MED-PC IV software (Med-associates).

Variable number lever-pressing task. After the house light was turned on, mice could initiate a trial by poking their noses into the tone site for 0.3 s. The 0.3-s nose poke triggered a speaker to generate a 2-s tone, after which a retractable lever was presented. The number of lever-presses required was randomly chosen as 8, 16, 32, 64, and infinity (reward omission) during each trial. After mice pressed the lever the required number of times, the lever was withdrawn, and 1 s after lever withdrawal, a food pellet was delivered to the reward site. Alternatively, mice could abandon the trial with a 0.3-s nose poke to the tone site. After reward delivery or abandonment of the trial, a 15-s inter-trial interval was inserted, which was indicated by turning off the house light. After the 15-s inter-trial interval, mice could initiate the next trial.

One session consisted of 53 trials (5 trials x 2 photostimulation conditions x 5 press number conditions + 3 trials with different photostimulation and press conditions). Two sessions were performed on a given testing day. Before testing sessions commenced, mice were trained to perform the lever-pressing task using the following schedule. Training took approximately 3 weeks.

All training sessions were performed either until mice earned 100 food pellets or

until 2 h, whichever came first. In order to habituate mice to the behavioral apparatus, they were first trained to poke their noses into the reward site to obtain a food pellet. Then they were trained to press a lever once to acquire a food pellet. Once mice could get more than 80 food pellets, they were trained to press a lever 3 times to obtain a food pellet. While the number of lever-presses required was progressively increased from 3, 5, 7, 10, 16, and 32 times, mice were trained that the lever was presented after a tone was generated. After the association between tone and lever presentation was established, mice were trained to poke their noses into the tone site to generate a tone for lever presentation. Training was completed when mice could get more than 80 rewards in a training session, during which they were required to press the lever 32 times after initiating a trial by nose pokes.

Tone-food waiting task. I used the same behavioral task as reported in previous studies [124, 128]. The same behavioral apparatus with a retractable lever was used for this task. In this task, mice could initiate a trial with a 0.3-s nose poke to the tone site, which triggered a 0.5-s tone. After hearing the tone, mice were required to continue poking their noses into the reward site. The required duration of the nose poke was randomly chosen as 2, 6, 10 s, and infinity (reward omission) during each trial. Once mice could wait for the required time, a food pellet was delivered to the reward site. For the optogenetic activation experiment, mice could initiate the next trial just after a reward delivery or after leaving the reward site. Because the suppression efficacy of ArchT decreases if sufficiently long intervals are not taken [113], for the photoinhibition experiment, the house light was turned off for 30 s after the end of a trial, and the next trial could be initiated once the house light was turned on again. One session consisted of 43 trials (5 trials x 2 photostimulation conditions x 4 delays + 3 trials with different photostimulation and delay conditions). Three sessions were performed on a testing day. Before the testing sessions, mice were trained for 2 hours, five days per week, and it took 2 weeks or less for mice to learn the task.

Surgical procedure for optic probe implantation. After training, a craniotomy was performed to implant an optic probe (400 µm diameter, 0.48 NA, 5 mm length, Doric) above the DRN. Mice were anesthetized with isoflurane (3% for induction and 1-1.5% during surgery). Mice were placed on a stereotaxic stage and their heads were fixed with ear bars. Then the skull was exposed with a blade, and a hole was made with a drill. Once the brain was exposed through the hole, the dura was removed using the tip of a needle, and the optic probe was lowered above the DRN through the hole (from bregma: posterior, -4.6mm; lateral, 0 mm; ventral, -2.6 mm). Light-sensitive adhesive and dental cement was applied to the skull to fix the implanted optic probe. Mice were placed back in their home cages for recovery. At least one week after the surgery, I started to retrain the mice for the behavioral tasks, and then commenced testing sessions.

Photostimulation protocol. During the testing sessions, 470-nm blue or 590-nm yellow light stimulation was given, generated by an LED light source (Doric Lenses). Timing of stimulation was determined by TTL pulses controlled by MED-PC IV software.

For the photoactivation experiment, the light intensities of blue and yellow light at the tip of the optical fibers were 1.6-2.0 mW and 1.1-2.0 mW, respectively. In both

tasks, in half the trials (selected at random), blue light stimulation was applied for activation, and in the other half of the trials, yellow light stimulation was given as a control. In the lever-pressing task, a 0.8-s blue/yellow light pulse was given when the mice started to press the lever and repeated at 20-s intervals. At the end of a trial, either when the mice pressed the lever the required number of times or when they abandoned the trial, a 1-s yellow light pulse was given to reset photoactivation. In the waiting task, a 0.8-s blue/yellow light pulse was given when mice first poked their noses into the reward site and 1-s yellow light pulse was given at the end of the trial, either when the mice waited until the end of the required delay or when they left the reward site.

For the photoinhibition experiment, the same LED light source was used. Intensities of blue and yellow light at the tip of the optical fibers was 2.8-3 mW and 2.8-3.2 mW, respectively. Continuous yellow/blue light was applied from the onset of action until the end of a trial.

Histological confirmation of implantation site. After the behavioral tests, mice were deeply anesthetized with 100 mg/kg sodium pentobarbital i.p. and perfused with saline or PBS followed by 4% PFA/PB or 4% PFA/PBS. Brains were removed immediately after perfusion and immersed in fixative solution overnight. Then, 50 μm coronal slices were cut using a vibratome (VT1000S, Leica) and the implantation site of optic probes was confirmed, according to the mouse brain atlas[64].

Immunohistochemistry. Brain slices were incubated with primary antibodies for 2 nights. Slices were then rinsed with PBS and incubated with secondary antibodies for 2 nights. After incubation and rinsing, slices were mounted on slide glasses. Fluorescent images (Figure 3.6a) were acquired using spinning disc confocal microscopy (SD-OSR, Olympus). As primary antibodies, I used anti-Tph (1:250, sheep polyclonal, Merck Millipore, AB1541) and anti-GFP (1:500, chicken polyclonal, Abcam, ab13970) as markers for 5-HT neurons and ChR2-EYFP or ArchT-EGFP neurons, respectively. For secondary antibodies, anti-sheep and anti-chicken antibodies conjugated with Alexa flour 594, and 488, respectively, were used. Antibodies were diluted in staining buffer containing 10 mM HEPES, 20 mM NaCl, and 10% Triton X-100. The pH of the staining buffer was adjusted to 7.4 in advance.

Behavioral parameters and statistical analysis. In the lever-pressing task, the success trial rate for 8-, 16-, 32-, and 64-press trials was calculated by dividing the number of rewarded trials by the total number of trials. In omission trials, the number of lever-presses, the time spent lever-pressing, which was defined as the time elapsed from the first lever-press to the last, and the time to abandon an omission trial, which was defined as the duration between the last lever-press and a nose poke to terminate the trial, were measured. To examine action vigor, inter-press intervals (IPIs), intervals between successive lever-presses, were measured. IPIs longer than 5 s were defined as long IPIs and below 5 s as short IPIs. In the waiting task, the time spent maintaining a nose poke was measured in omission trials. Behavioral parameters were calculated using custom-written programs in MATLAB.

Statistical tests were selected based on whether the data satisfied normality and homogeneity of variance, assessed with the Shapiro-Wilk test and Levene test, respectively. If the data satisfied these assumptions, I used paired t-tests for within-group

comparisons and unpaired t-tests for group comparisons. If not, I used Wilcoxon signed-rank tests for within-subject comparisons and Mann Whitney U-tests for group comparisons. Statistical analysis was performed using Python. Two-way ANOVA was performed using SPSS.

3.3 Results

3.3.1 Photoactivation of DRN 5-HT neurons and behavioral tasks.

I used eight Tph2-ChR2(C128S)-EYFP bi-transgenic mice (hereafter, referred to as ChR2 mice) to selectively activate DRN 5-HT neurons following a pulse of blue light, as in previous studies [124, 128]. Five Tph2-tTA mice were used as the control group (Hereafter, referred to as control mice). These mice were trained to perform two different operant conditioning tasks: a repeated lever-pressing task and a stationary waiting task.

Behavioral tasks: In the lever-pressing task (Fig. 3.1a), a mouse was required to press a lever multiple times to obtain a food pellet after its voluntary 0.3-s nose poke to the tone site. A trial was ended either when the mouse successfully obtained a reward by pressing the lever the required number of times or abandoned the trial with a 0.3-s nose poke to the tone site. At the first lever-press, a 0.8-s blue or yellow light pulse was applied and repeated at 20-s intervals. At the end of the trial, which was defined by either a reward delivery or a nose poke to the tone site to abandon the trial, a 1-s yellow light pulse was applied to stop activation (Fig. 3.1b).

I used the stationary waiting task developed in previous studies [124, 128] (Fig. 3.1c). In this task, a mouse was required to keep poking its nose in the reward site after a voluntary 0.3-s nose poke to the tone site. The delay was randomly chosen as 2 s, 6 s, and 10 s, and infinity (reward omission) in every trial. A 0.8-s blue or yellow light pulse was randomly applied to induce activation or no activation of DRN 5-HT neurons at the onset of nose poking at the reward site and a 1-s yellow-light pulse was applied at the end of the trial, which was defined by either reward delivery or leaving the reward site (Fig.3.1d).

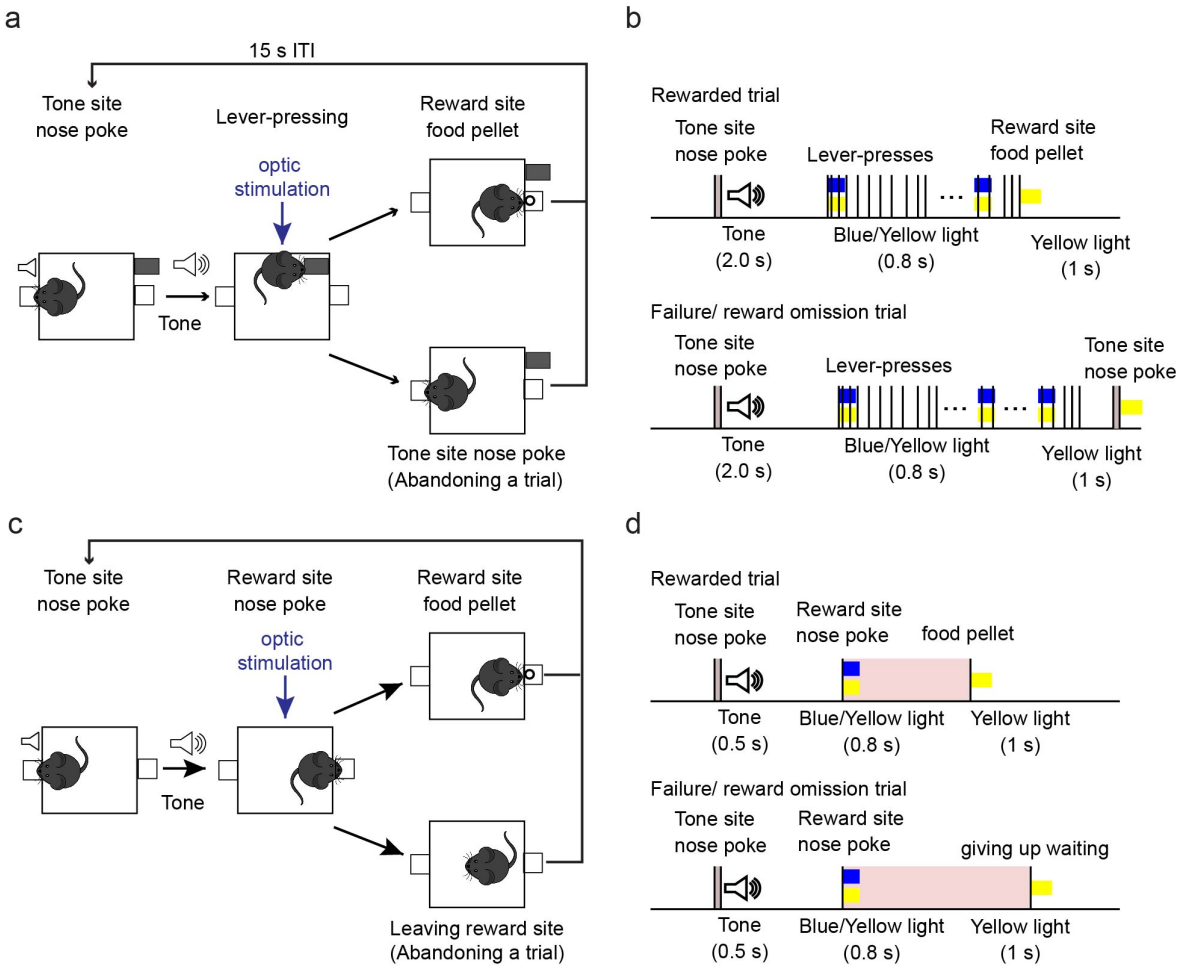

Figure 3.1: Behavioral tasks and timing of optogenetic activation. a: Schematic drawing of the lever-pressing task. b: Time sequence of rewarded and failure/reward omission trials with optic stimulation during the lever-pressing task. c: Schematic drawing of the stationary waiting task. d: Time sequence of rewarded and failure/reward omission trials with optic stimulation during the waiting task.

Histological Confirmation: In order to optically stimulate DRN 5-HT neurons, optic probes were implanted above the DRN [64]. Six of the eight ChR2 mice were sacrificed to confirm the implantation site of the optic probes. Although the site varied along the anterior-posterior axis, all probes examined were located above the DRN (Fig.3.2).

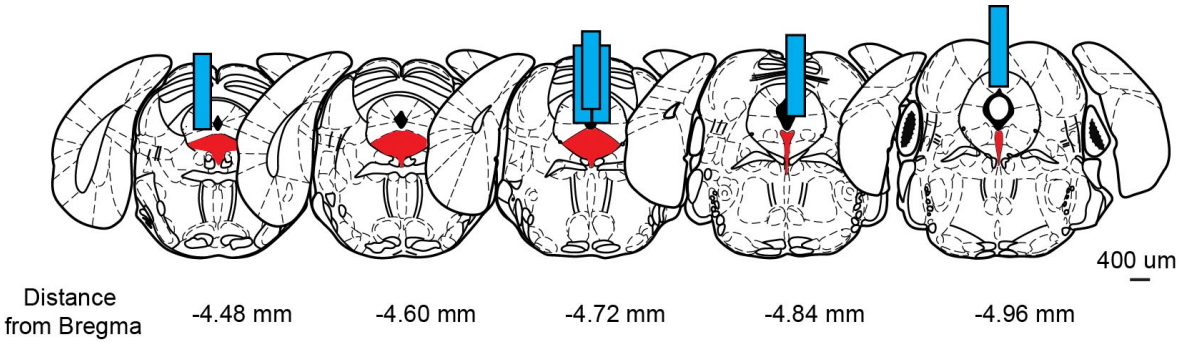

Figure 3.2: The implantation site of optic probes in ChR2 mice. Coronal views of the mouse brain are adapted from [64]. Blue rectangles indicate tracks of implanted optic fibers. Red-filled areas indicate the DRN.

3.3.2 Activation of DRN 5-HT neurons prolonged stationary waiting for future rewards.

To confirm the effectiveness of photoactivation, I examined whether the photoactivation protocol used here affected waiting behaviors for delayed rewards of 2 s, 6 s, 10 s or infinity (reward omission). The waiting duration, the duration of maintaining a nose poke, was measured in each omission trial of the waiting task (Fig. 3.3a). Optogenetic activation with blue light significantly increased waiting duration during omission trials in ChR2 mice (yellow vs. blue trials: 15.0 ± 0.37 s vs. 17.4 ± 0.49 s (hereafter, mean ± SEM is shown otherwise indicated); p = 0.00040, paired t-test; Fig. 3.3b). The change rate was significant compared to that of control mice (Control vs. ChR2 mice: -0.012 ± 0.0087 vs. 0.16 ± 0.027; p = 0.00042 , unpaired t-test; Fig. 3.3b-d). This result was consistent with previous studies using the same behavioral task [124, 128] and confirmed that the photoactivation administered to these mice was sufficient to induce behavioral changes.

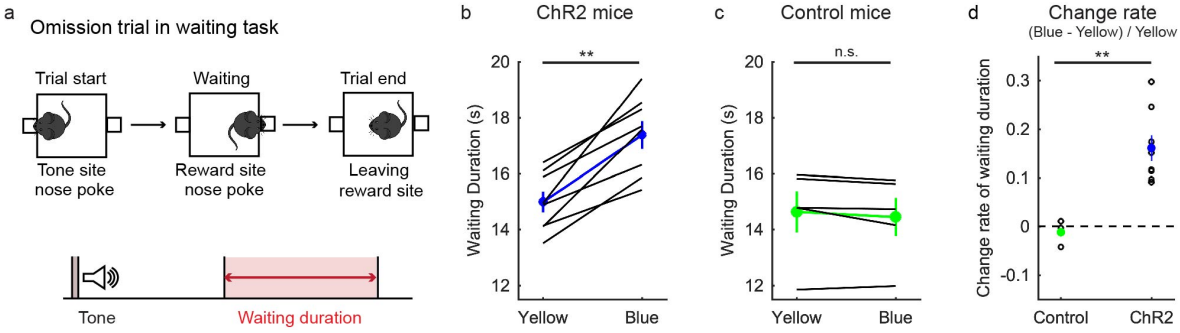

Figure 3.3: Activation of DRN 5-HT neurons prolonged stationary waiting for future rewards. a: The definition of waiting duration. b, c: Waiting duration in an omission trial in ChR2 ($n = 8$ mice) and control ($n = 5$ mice) mice. Blue and green dots indicate the mean across ChR2 and control mice, respectively. ** indicates $p < 0.01$ by paired t-test. d: Change rate in control ($n = 5$ mice) and ChR2 ($n = 8$ mice) mice. Green- and blue-filled circles indicate the mean across control ($n = 5$ mice) and ChR2 ($n = 8$ mice) mice respectively. ** indicates $p < 0.01$, unpaired t-test. The error bars represent the SEM in all graphs.

3.3.3 Activation of DRN 5-HT neurons neither enhanced nor suppressed persistence in motor actions.

Persistence in motor actions

In order to examine whether photoactivation of DRN 5-HT neurons affects persistence in motor actions for future rewards, I analyzed the successful trial rate, the duration, the number of lever presses in omission trials, and the time spent in abandoning an omission trial in the lever-pressing task.

Successful trial rate: I first calculated the percentage of successfully rewarded trials in 8-, 16-, 32-, and 64-press trials. In 8-, 16-, and 32-press trials, mice successfully obtained rewards with almost 100% of the time. In 64-press trials, the successful trial rate decreased, but was not significantly different between blue light and yellow light stimulation (yellow vs. blue 64-press trials: 95.02 ± 1.73% vs. 94.55 ± 3.23%; p = 0.83, paired t-test; Fig. 3.4a).

Time spent pressing the lever: To quantify how long mice could sustain actions for delayed rewards, I next measured the time spent pressing the lever, the duration from the first lever-press to the last lever-press, in omission trials (Fig. 3.4c(i)). Interestingly, the mice spent more than three times longer pressing the lever (48.68 ± 3.09 s with yellow light, Fig. 3.4d(i)) than they spent in stationary waiting (15.0 ± 0.37 s, Fig. 3.3b) for the same reward, showing that mice can tolerate longer delays while they are actively engaged in doing something, as opposed to waiting inactively. However, the time spent lever-pressing in an omission trial was not significantly different between trials with activation and those without activation (yellow vs. blue trials: 48.68 ± 3.09 vs. 48.86 ± 2.27; p = 0.92, paired t-test; Fig. 3.4d(i)). The change rate was not significantly different from that of control mice (Control vs. ChR2 mice: 0.0048 ± 0.0052 vs. 0.016 ± 0.042; p = 0.88, unpaired t-test; Fig. 3.4d-f(i)).

The number of lever-presses in omission trials: To quantify how persistently mice sustained motor actions for future rewards, I measured the number of lever-presses in omission trials (Fig. 3.4c(ii)). The number of lever-presses in omission trials with activation was not significantly different than that without activation (yellow vs. blue trials: 111.50 ± 3.09 vs. 108.63 ± 3.74; p = 0.38, paired t-test; 3.4d(ii)). The change rate was not significantly different from that of control mice (Control vs. ChR2 mice: -0.016 ± 0.041 vs. -0.024 ± 0.029; p = 0.87 , unpaired t-test; Fig. 3.4d-f(ii)).

Time needed to abandon a trial: I next measured the time from the last lever-press to a nose poke in the tone site to abandon an omission trial, which could indicate how ambivalent mice were about abandoning the present trial (Fig. 3.4c(iii)). In ChR2 mice, optogenetic activation did not significantly change the time spent to abandon a trial (yellow vs. blue trials: 15.41 ± 3.53 s vs. 19.03 ± 6.37 s; p = 0.23, Wilcoxon signed-rank test; Fig. 3.4d(iii)) and the change rate was not significantly different from that of control mice (Control vs. ChR2 mice: -0.059 ± 0.096 vs. 0.13 ± 0.090; p = 0.19 , unpaired t-test; Fig. 3.4d-f(iii)). These results indicate that DRN 5-HT activation neither enhanced nor suppressed sustained motor actions for future rewards.

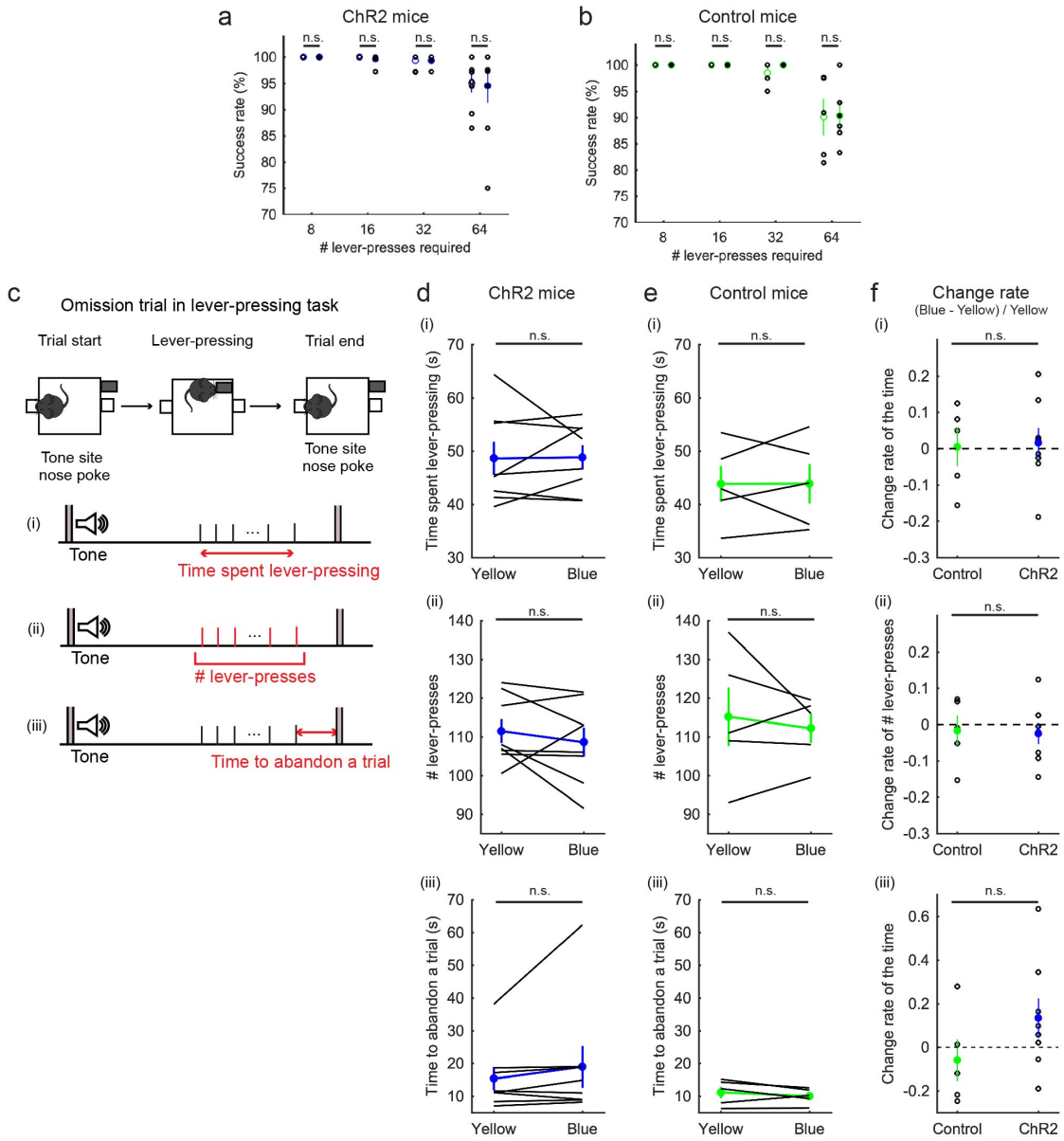

Figure 3.4: Activation of DRN 5-HT neurons did not change persistence in motor actions for future rewards. a, b: Successful trial rates in ChR2 (n = 8 mice) and control (n = 5 mice) mice. Open and filled circles indicate the means of blue and yellow light trials, respectively in ChR2 (blue) and control (green) mice. n.s. indicates not significant (p > 0.05) by paired t-test. c: The definition of behavioral measures for action persistence. d, e: Behavioral parameters in ChR2 (n = 8 mice) and control (n = 5 mice) mice. Blue and green dots indicate the means of ChR2 and control mice data, respectively. n.s. indicates not significant (p > 0.05) by paired t-test in d(i) and d(ii) and by Wilcoxon signed-rank test in d(iii). f: Change rate of behavioral parameters in control (n = 5 mice) and ChR2 (n = 8 mice) mice. Green- and blue-filled circles indicate the mean across control and ChR2 mice, respectively. n.s. indicates no significance (p > 0.05) by unpaired t-test. The error bars represent the SEM in all graphs.

Action vigor

In order to examine how optogenetic activation of DRN 5-HT neurons affects the vigor or speed of sustained motor actions, I measured inter-press intervals (IPIs), the intervals between successive lever-presses. Through behavioral observation, I found that mice usually pressed the lever continuously, but sometimes showed behaviors irrelevant to lever-pressing, such as pausing, resting, or exploring, especially in omission trials. To be specific, most IPIs were < 5 s, but some were longer (Fig. 3.5a). Therefore, I defined IPIs < 5 s as short IPIs, which represent vigorous lever-pressing behaviors, and IPIs ≥ 5 s as long IPIs, which mainly represent other behaviors, and examined the effect of the photoactivation on each type of IPI.

Long IPIs: Long IPIs were analyzed only in omission trials, because they were rarely found in trials requiring 8 or more presses. In ChR2 mice, there was no significant difference in long IPIs between blue light and yellow light stimulation. (yellow vs. blue trials: 16.94 ± 2.65 s vs. 19.80 ± 3.70 s; p = 0.23, Wilcoxon signed-rank test; Fig. 3.5b). The change rate was not significantly different from that of control mice (Control vs. ChR2 mice: 0.17 ± 0.096 vs. 0.17 ± 0.11; p = 0.998, unpaired t-test; Fig. 3.5b-d)).

Short IPIs: In analyzing short IPIs, I first calculated the average short IPI in a trial and analyzed the median across trials for each mouse. I analyzed data of ChR2 and control mice with a two-way repeated measures ANOVA (Fig. 3.5e for ChR2 and 3.5f for control mice). There were significant main effects of light (two levels within-subject factors; yellow and blue, $F(1,7) = 20.14$, $P < 0.004$) and press (five levels within-subject factors; 8-press, 16-press, 32-press, 64-press, and omission, $F(4,28) = 26.39$, $P < 10^{-8}$). However, there was no significant main effect of interaction (light x press, $F(4,28) = 0.68$, $P = 0.62$). On the other hand, in control mice there was a significant main effect of press (five levels within-subject factors; 8-press, 16-press, 32-press, 64-press, and omission, $F(4,16) = 19.7$, $P < 0.04$), but there was no significant main effect of light ($F(1,4) = 0.148$, $P < 0.72$).

Short IPIs at different press timing: Lottem et al. [104] showed that the vigorousness of actions reflects latent variables of decision making. Accumulated evidence of reward omission reduced response vigor [104]. In the lever-pressing task, because mice could accumulate evidence of reward omission as they pressed the lever more times, the vigorousness of lever-presses dynamically changed as a function of the number of lever-presses. To confirm this point, I calculated the averages of short IPIs before 8 lever-presses, from 8 to 16 lever-presses, from 16 to 32 lever-presses, from 32 to 64 lever-presses, and after 64 lever-presses. Short IPIs increased as mice experienced more lever-presses both in ChR2 and control mice and both in blue and yellow light stimulation (Fig. 3.5g for ChR2 and 3.5h for control mice). This result suggests short IPIs may reflect the dynamic change of the likelihood of reward or increased probability of reward omission.

I analyzed data of ChR2 and control mice with a two-way repeated measures ANOVA. In ChR2 mice, there were significant main effects of light (two levels within-subject factors; yellow and blue, $F(1,7) = 20.92$, $P < 0.004$) and press (five levels within-subject factors; before 8 presses, 8-16 presses, 16-32 presses, 32-64 presses, after

64 presses, $F(4,28) = 22.93$, $P = 0.00044$). However, there was no significant main effect of interaction (light x press, $F(4,28) = 3.31$, $P = 0.073$). There was a significant simple main effect of light ($F(1,7) = 20.14$, $P < 0.004$). On the other hand, in control mice there was a significant main effect of press (five levels within-subject factors; before 8 presses, 8-16 presses, 16-32 presses, 32-64 presses, after 64 presses, $F(4,28) = 22.93$, $P < 10^{-5}$), but there was no significant main effect of light ($F(1,4) = 0.134$, $P = 0.73$). These results showed that activation of DRN 5-HT neurons slowed the speed of active lever-pressing although the same activation did not affect persistence of sustained motor actions. Taken together with the results of the waiting task, these results show that the effect of activation of DRN 5-HT neurons on sustained actions for future rewards differs between active motor actions and stationary waiting.

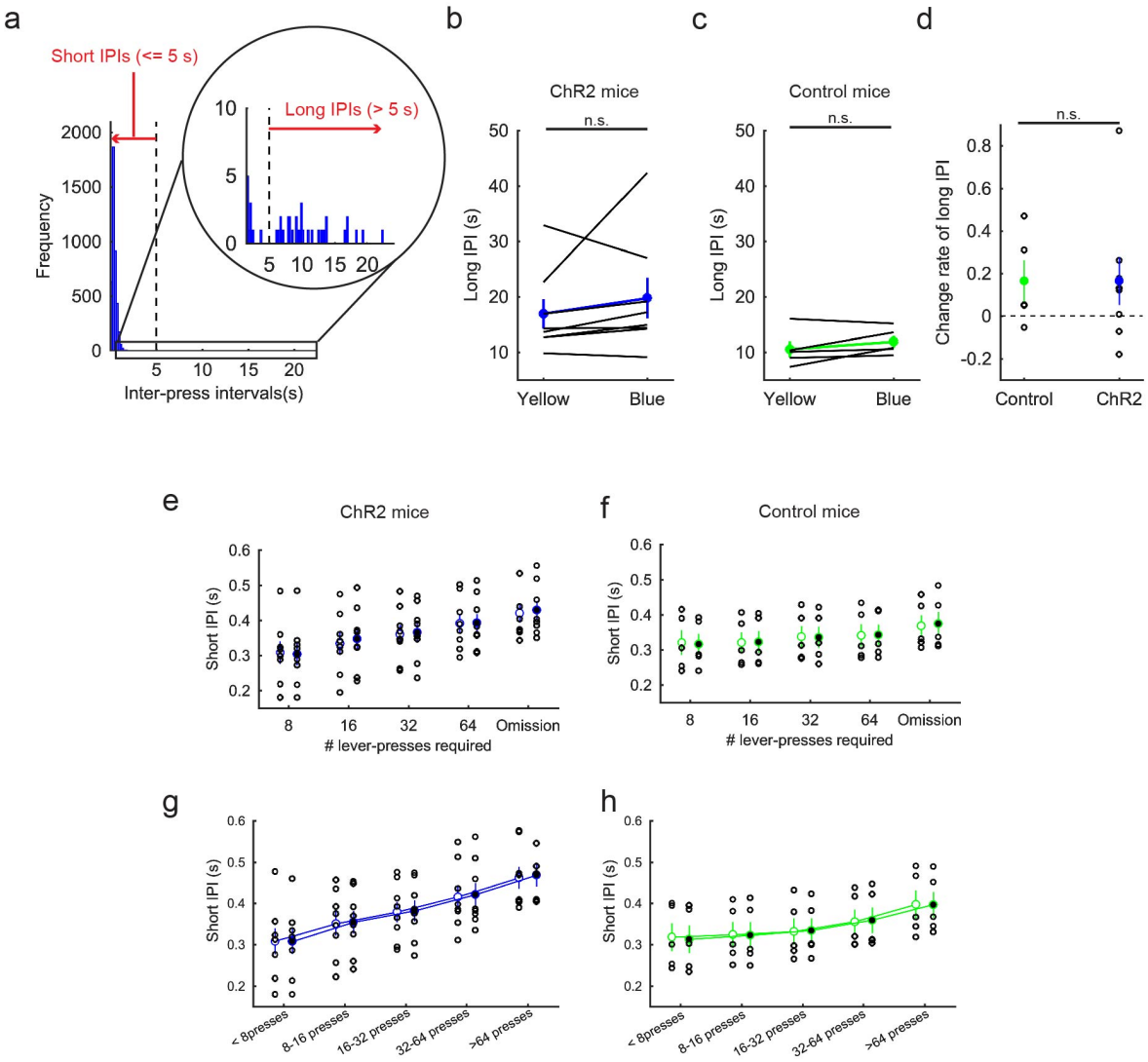

Figure 3.5: Activation of DRN 5-HT neurons did not change action vigor in the lever-pressing task. a: The definition of long and short IPIs. b, c: Long IPIs in omission trials in ChR2 ($n = 8$ mice) and control ($n = 5$ mice) mice. Blue and green dots indicate the mean across ChR2 and control mice, respectively. n.s. indicates not significant ($p > 0.05$) by paired t-test. d: Change rate of long IPIs in control ($n = 5$ mice) and ChR2 ($n = 8$ mice) mice. Green- and blue-filled circles indicate the mean across control and ChR2 mice, respectively. n.s. indicates not significant ($p > 0.05$) by unpaired t-test. e, f: Short IPIs in ChR2 ($n = 8$ mice) and control ($n = 5$ mice) mice. Open and filled circles indicate the mean in blue and yellow light trials, respectively in ChR2 (blue) and control (green) mice. g, h: Short IPIs at different press timing in ChR2($n = 8$ mice) and control ($n = 5$ mice) mice. Open and filled circles indicate the mean in blue and yellow light trials, respectively in ChR2 (blue) and control (green) mice. The error bars represent the SEM in all graphs.

3.3.4 Photoinhibition of DRN 5-HT neurons and behavioral tasks.

A previous study showed that a subset of putative DRN 5-HT neurons increased their neural activity during behavioral activation, such as locomotion, changing direction, and approaching/withdrawal behaviors [150]. A possibility given this observation is that lever-pressing behavior itself increases neural activity of DRN 5-HT neurons, such that optogenetic activation did not induce additional effects. Therefore, I examined the effect of optogenetic inhibition of DRN 5-HT neurons on action maintenance.

Histological Confirmation: To selectively inhibit DRN 5-HT neurons, I prepared Tph2-ArchT-EGFP bi-transgenic mice (hereafter, referred to as ArchT mice). These transgenic mice selectively express ArchT, a light-sensitive proton pump, in 5-HT neurons. Under yellow light, ArchT induces efflux of H^+ and inhibits neural activities. To confirm selective expression of ArchT, I performed a histological experiment to compare cells expressing EGFP and 5-HT neurons identified by immunohistochemistry of Tph2 in three ArchT mice that were not used for the optogenetic inhibition experiment. In nine slices from the three ArchT mice, 1058 Tph^+ cells were found. Among them, 72.5% of the cells were also ArchT-$EGFP^+$. On the other hand, there were only three Tph^- but ArchT-$EGFP^+$ cells, suggesting the Tph2-ArchT bi-transgenic mice selectively expressed ArchT in 5-HT neurons (Fig. 3.6a). I also confirmed the implantation site of optic probes. Optic probes were implanted above the DRN in ArchT mice (Fig. 3.6b).

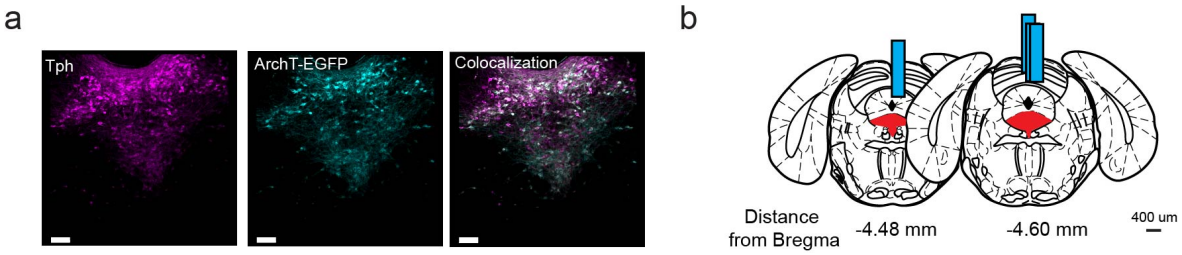

Figure 3.6: Immunohistochemistry and the implantation site of optic probes in ArchT mice. a: Fluorescence images from ArchT mice. These images indicate expression of Tph (Left), ArchT-EGFP (Middle), and co-localization of the two signals (Right). Scale bars indicate 100 µm. b: The implantation site of optic probes in ArchT mice. Coronal views of mouse brain are adapted [64]. Blue rectangles indicate tracks of the implanted optic probes. Red filled areas indicate the DRN.

Behavioral tasks: In this experiment, I slightly modified the task design of the stationary waiting task in that 30-s inter-trial intervals were inserted after each trial (Fig. 3.7a for the lever-press and Fig. 3.7c for the stationary waiting task). In order to optogenetically inhibit DRN 5-HT neurons, continuous yellow light was applied from the onset of action until the end of the trial (Fig. 3.7b for the lever-press and Fig. 3.7d for the stationary waiting task). Continuous blue light stimulation was used for controls.

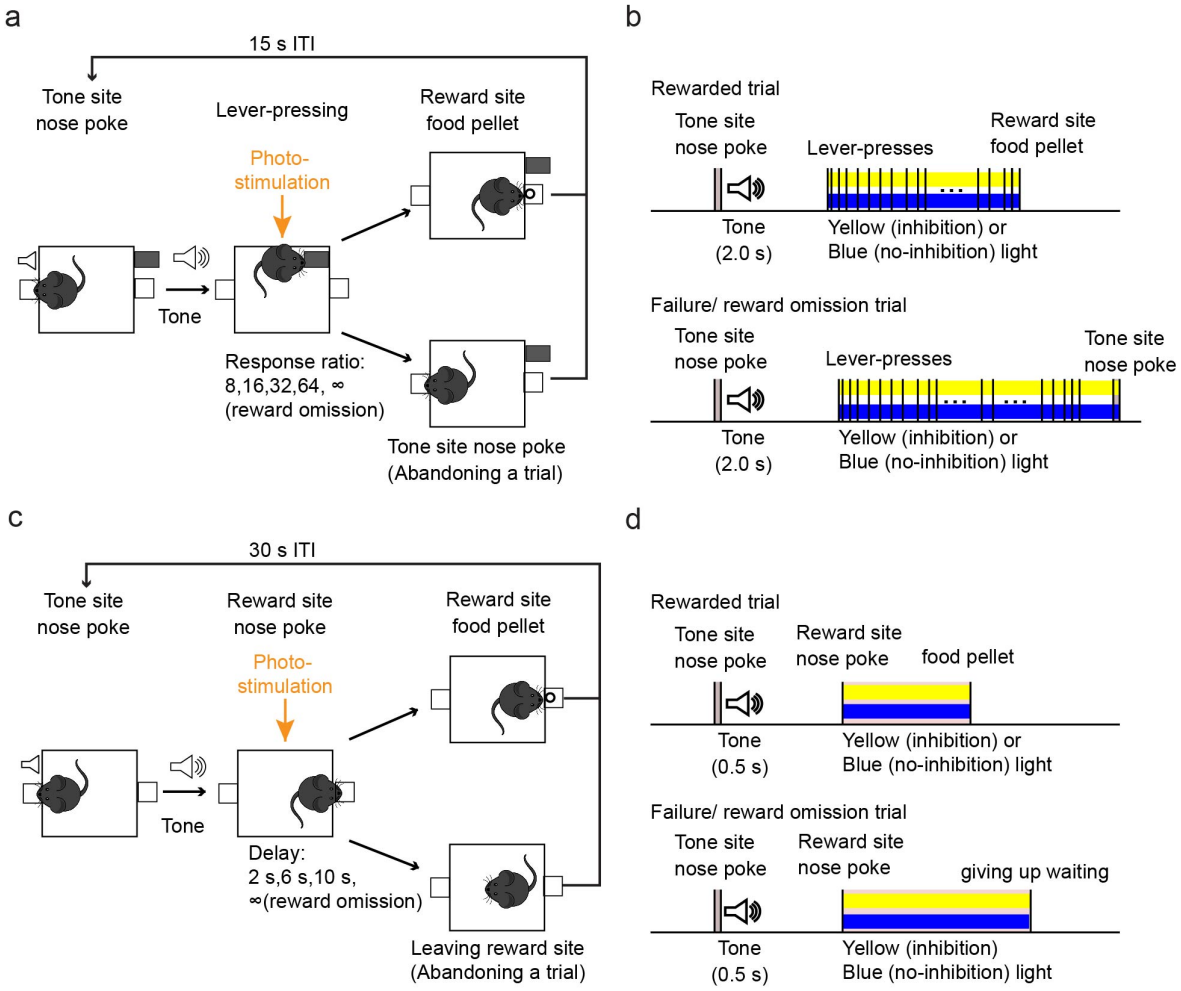

Figure 3.7: Behavioral tasks and timing of optogenetic inhibition. a: Schematic drawing of the lever-pressing task. b: Time sequence of rewarded and failure/reward omission trials with optic stimulation during the lever-pressing task. c: Schematic drawing of the stationary waiting task. d: Time sequence of rewarded and failure/reward omission trials with optic stimulation during the waiting task.

3.3.5 Inhibition of DRN 5-HT neurons shortened stationary waiting for future rewards.

I first trained four ArchT mice in the stationary waiting task and tested the effect of optogenetic inhibition of DRN 5-HT neurons. A previous study showed that chemical inhibition of DRN 5-HT neurons increased premature abandonment in a delayed reward task [127]. Therefore, I predicted that optogenetic inhibition of DRN 5-HT neurons would decrease waiting duration in omission trials or increase wait errors in reward trials in the stationary waiting task. I tested the photoinhibition effect in the condition in which the reward delay was randomly chosen from 2, 6, and 10 s. I analyzed successful trial rates using two-way repeated measures ANOVA (Fig. 3.8a). There were no significant main effects of light (two levels within-subject factors; yellow and blue, $F(1,3) = 1.13$, $P = 0.37$) and delay (three levels within-subject factors; 2 s, 6 s, 10 s, $F(2,6) = 0.19$, $P = 0.83$).

On the other hand, there was a significant decrease in the waiting duration in omission trials during optogenetic inhibition (blue vs. yellow trials: 17.08 ± 0.55 vs. 15.50 ± 0.39; $p = 0.0350$, paired t-test, $n = 4$ mice; Fig. 3.8b). These results confirmed the effectiveness of the optogenetic inhibition protocol and also the causal relationship between decreased DRN 5-HT neural activity and impaired waiting for delayed rewards.

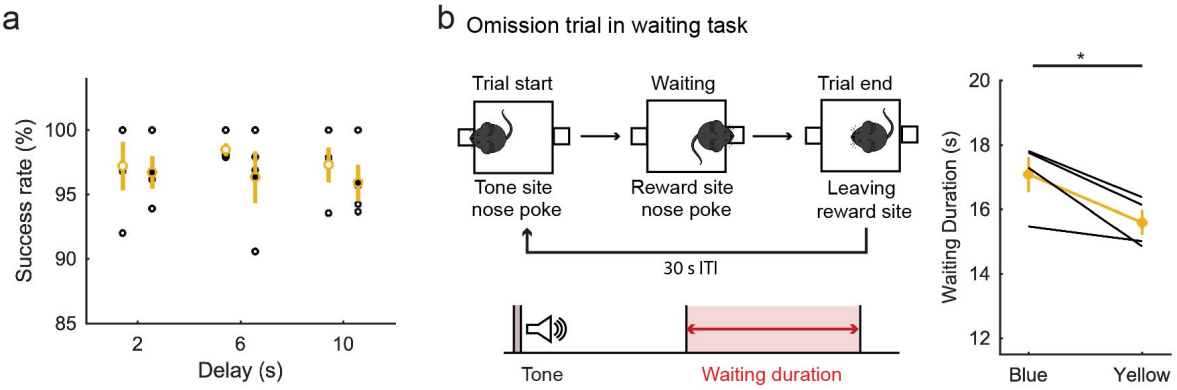

Figure 3.8: Inhibition of DRN 5-HT neurons shortened the stationary waiting periods for future rewards. a: The successful trial rates in rewarding trials in ArchT mice ($n = 4$ mice). Yellow open and filled circles indicate the mean of ArchT mice in blue and yellow light trials, respectively. b: Waiting duration in omission trials in ArchT mice ($n = 4$ mice). Yellow dots indicate the mean across ArchT mice ($n = 4$ mice). * indicates $p < 0.05$ by paired t-test. Error bars represent the SEM in all graphs.

3.3.6 Inhibition of DRN 5-HT neurons neither enhanced nor suppressed sustained motor actions

Persistence in motor action

In order to examine whether optogenetic inhibition of DRN 5-HT neurons affects sustained motor actions for future rewards, I analyzed the successful trial rate, the duration, the number of lever presses in omission trials, and the time spent abandoning a trial in the lever-pressing task, as in the optogenetic activation experiment.

Successful trial rate: All mice were able to perform 8-, 16-, and 32-press trials without failure in both in blue and yellow light stimulation trials. In 64-press trials, the successful trial rate decreased compared with other types of trials, but there was no significant difference between blue and yellow light trials (blue vs. yellow 64 press trials: 97.54 ± 1.77% vs. 98.21 ± 1.14% ; p = 0.81, paired t-test; Fig. 3.9a).

Time spent pressing the lever: I measured the time that mice spent pressing the lever, the time from the first to the last lever-press, in omission trials. The time spent lever-pressing in omission trials (44.62 ± 6.64 s with blue light, Fig. 3.9b(i)) was much longer than that for stationary waiting in omission trials (17.08 ± 0.55 s with blue light, Fig. 3.8b), as in the optogenetic activation experiment, but did not differ significantly between trials with and without optogenetic inhibition (blue vs. yellow trials: 44.62 ± 6.64 s vs. 44.93 ± 7.02 s; p = 0.77, paired t-test; Fig. 3.9b(i)).

The number of lever-presses in omission trials: I next measured the number of lever-presses in omission trials. That number did not differ significantly between trials with and without optogenetic inhibition (blue vs. yellow trials: 107.63 ± 6.28 vs. 111.00 ± 4.88; p = 0.14, paired t-test, n = 4 mice; Fig. 3.9b(ii)).

The time needed to abandon a trial: I also measured the time from the last lever-press to a nose poke in the tone site to abandon an omission trial. Optogenetic inhibition did not significantly change the time to abandon an omission trial (blue vs. yellow trials: 13.52 ± 3.69 s vs. 10.50 ± 2.07 s; p = 0.17, paired t-test; Fig. 3.9b(iii)).

These results indicate that DRN 5-HT inhibition neither enhanced nor suppressed persistence in motor actions, thereby resembling results of the photoactivation experiment.

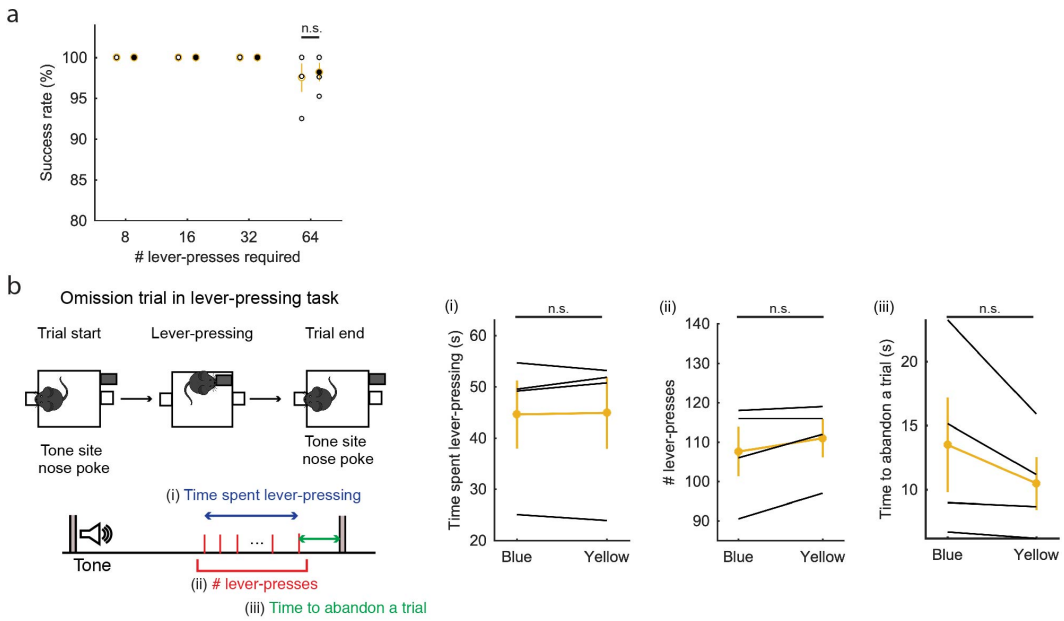

Figure 3.9: Inhibition of DRN 5-HT neurons neither enhanced nor suppressed persistence in motor actions for future rewards. a: The successful trial rates in rewarding trials in ArchT mice ($n = 4$ mice). Yellow open and filled circles indicate the mean among ArchT mice in blue and yellow light trials, respectively. n.s. indicates not significant ($p > 0.05$) by paired t-test. b: Behavioral parameters in an omission trial in ArchT mice ($n = 4$ mice). Yellow dots indicate the mean of ArchT mice. n.s. indicates not significant ($p > 0.05$) by paired t-test in all parameters. Error bars represent the SEM in all graphs.

Action vigor

To examine the effect of optogenetic inhibition on the vigor or speed of motor actions, I measured IPIs. As in the optogenetic activation experiment, I defined IPIs < 5 s as short IPIs, which represent vigorous lever-pressing behavior, and IPIs ≥ 5 s as long IPIs, and I examined the effect of optogenetic inhibition on each type of IPI (Fig. 3.10a).

Long IPIs: I first examined the effect of optogenetic inhibition on long IPIs in omission trials. Long IPIs did not change significantly as a result of optogenetic inhibition (blue vs. yellow trials: 19.82 ± 3.48 s vs 15.42 ± 1.27 s; p = 0.15, paired t-test; Fig. 3.10b).

Short IPIs: I then examined short IPIs in different types of trials. I analyzed data with a two-way repeated measures ANOVA (Fig. 3.10c). There were no significant main effects of light (two levels within-subject factors; yellow and blue, $F(1,3) = 9.63$, $P = 0.053$) and press (five levels within-subject factors; 8-press, 16-press, 32-press, 64-press, and omission, $F(4,12) = 0.045$, $P = 0.99$).

Short IPIs at different press timing: To examine how the speed of lever-presses dynamically changed, I measured short IPIs at different press timing, as in the photoactivation experiment. Compared to the photoactivation experiment, I did not find a clear trend that short IPIs increased as mice experienced more lever-presses. An increase in short IPIs at a larger number of lever-presses was found in two of the four ArchT mice, while the other two ArchT mice maintained the same speed during the lever-presses. I analyzed data with a two-way repeated measures ANOVA (Fig. 3.10d). There were no significant main effects of light (two levels within-subject factors; yellow and blue, $F(1,3) = 1.25$, $P = 0.36$) and press (five levels within-subject factors; before 8 presses, 8-16 presses, 16-32 presses, 32-64 presses, after 64 presses, $F(4,12) = 0.27$, $P = 0.89$).

These results showed that DRN 5-HT inhibition did not modulate the vigor of motor actions. Taken together with the results of the waiting task, these results suggest that involvement of DRN 5-HT in sustained actions for future rewards differs between physical activity and stationary waiting. DRN 5-HT neurons bidirectionally controlled stationary waiting, but they did not affect persistence in motor actions.

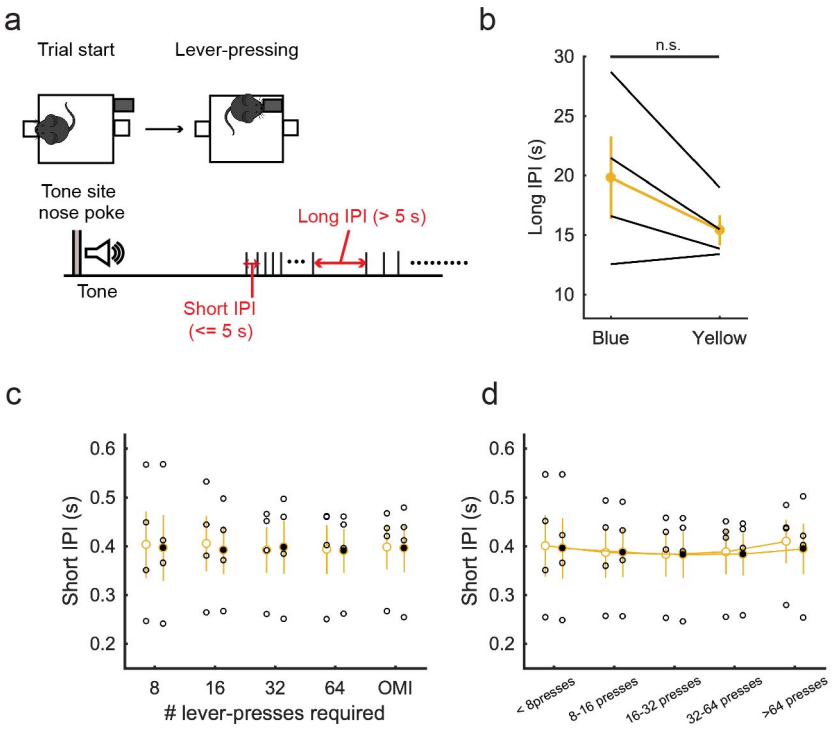

Figure 3.10: Inhibition of DRN 5-HT neurons did not change action vigor in the lever-pressing task. a: The definition of long and short IPIs. b: Long IPIs at omission trials in ArchT mice ($n = 4$ mice). Yellow dots indicate the mean of ArchT mice. n.s. indicates not significant ($p > 0.05$) by paired t-test. c: Short IPIs in ArchT mice ($n = 4$ mice). Yellow open and filled circles indicate the mean of ArchT mice in blue and yellow light trials, respectively. d: Short IPIs at different press timing in ArchT mice ($n = 4$ mice). Yellow open and filled circles indicate the mean across ArchT mice in blue and yellow light trials, respectively. Error bars represent the SEM in all graphs.

3.4 Discussion

In this study, I examined the role of DRN 5-HT neurons in sustaining motor actions for future rewards. I first showed that optogenetic activation of DRN 5-HT neurons prolonged waiting for future rewards, as in previous studies. On the other hand, the same stimulation neither enhanced nor suppressed the persistence of motor actions, while stimulation reduced response vigor. This result suggests that activation of DRN 5-HT neurons, which promotes patience to wait, does not modulate persistence in motor actions.

To further investigate the causal relationship between the activity of DRN 5-HT neurons and sustained motor actions, I optogenetically inhibited DRN 5-HT neurons during behavioral tasks. Optogenetic inhibition of DRN 5-HT neurons shortened the waiting duration, in opposition to its effect in the activation experiment. However, the same optogenetic inhibition did not cause behavioral changes in lever-pressing behavior. Taken together, the present study revealed differential regulation of sustained motor actions and inactive waiting for future rewards.

The first interesting behavioral observation was that animals spent more time pressing a lever than continuing to nose poke in omission trials. That animals tolerate longer delays to obtain the same reward by sustained active behavior than by stationary waiting is not compatible with the common notion that motor actions are costly. Previous psychological studies on self-control in humans showed that children who are engaged in behaviors to distract themselves, such as playing with toys or talking to others, can tolerate longer delays for rewards than those who did not [120, 121]. From these studies, it was proposed that diverting attention from the temporal feature of the task with activity reduces the effect of delay discounting [116]. In the current study, mice might be able to tolerate longer delays during the lever-pressing task due to actively sustained motor actions that shifts their attention away from the temporal feature of the task. In order to examine this, for example, we can train mice to the task in which mice need to run or not to run in a wheel for required delay to obtain a reward and compare the success rate in running and no-running trial and waiting duration in reward omission trials.

Based on the hypothesis that 5-HT controls temporal discounting of future rewards, we originally expected that optogenetic manipulation of DRN 5-HT neural activity would modulate sustained motor actions for delayed rewards in much the same way during stationary waiting. However, these experimental results suggest that the role of DRN 5-HT neurons in sustained motor actions cannot be fully explained by the discount factor hypothesis. Another possible interpretation of how activation of DRN 5-HT neurons prolong waiting is based on the behavioral inhibition hypothesis, which suggests that increased 5-HT transmission shifts animal behavior toward inaction [167]. If this hypothesis can account for 5-HT regulation of adaptive behaviors, it might be expected that activation of DRN 5-HT neurons would suppress lever-pressing behavior and vice versa. However, in the present study, optogenetic stimulation of DRN 5-HT neurons neither enhanced nor suppressed persistent lever-pressing, suggesting that behavioral inhibition cannot explain the results of the present study.

The differential optogenetic effect in the present study may reflect distinctive neural substrates for delay-based (waiting) and effort-based (lever-pressing) motivated behav-

iors. A previous study showed that pharmacological reduction of systemic 5-HT level decreased the tendency of rats to favor immediate smaller rewards over delayed larger rewards in a delay-based choice task, but did not change their tendency to choose smaller rewards after climbing a low barrier rather than larger rewards after climbing a higher barrier in an effort-based choice task [47]. Although sustained actions may be regulated differently than choice behaviors, our results are consistent with that behavioral study. Also, a recent fiber photometry study showed that DRN 5-HT neural activity is not increased while mice are pressing a lever to obtain rewards [207]. Rather, the study showed that 5-HT neurons in the MRN show increased neural activity during lever-pressing. How DRN 5-HT neurons regulate sustained actions differently has not been examined behaviorally, and our study adds behavioral evidence suggesting different neural substrates between sustained motor action and inactive waiting for future rewards.

A recent study showed that DRN 5-HT neurons are anatomically divided into at least two subtypes, cortex-projecting and subcortical-projecting neurons, and that these subtypes show different responses to rewarding and punishing stimuli. Moreover, they have different behavioral roles [152]. Therefore, it is also possible that DRN 5-HT projections to different brain regions have different behavioral functions. In our study, we stimulated all DRN 5-HT neurons, which may have obscured the existence of functionally different DRN 5-HT projections on motor action for future rewards. Future experiments to stimulate axon terminals of DRN 5-HT projections to specific brain regions [125] should further clarify the roles of DRN 5-HT neurons in sustained motor actions.

Another possible reason why optogenetic inhibition of DRN 5-HT neurons did not have any effect on lever-pressing behavior may be the order of behavioral experiments. In this study, we first tested the effect of optogenetic inhibition in the stationary waiting task. After the waiting task, we trained mice for the lever-pressing task. A previous study showed that excessively long optogenetic stimulation, i.e., 2 h, caused cell death, possibly due to extreme alkalinization of intracellular pH [178]. Although the present study did not use prolonged optogenetic stimulation as in the aforementioned study, repeated optogenetic stimulation over multiple days may have reduced the efficacy of such stimulation in later testing sessions. It is necessary to reverse the order of the behavioral tasks to exclude the influence of possible photodamage.

Although the present study shows no effect on action persistence for future rewards and but an effect of slowing down response vigor, this result conflicts with the previous study, which showed that optogenetic activation of DRN 5-HT neurons promotes active persistence and increased response vigor in a task in which active nose poking was required to receive water rewards [104]. Although the previous study differs from ours in the type of reward (water vs. food) and in actions associated with rewards (repeated nose pokes vs. lever-pressing). Other studies also discussed possible reasons of different optogenetic effect by studies such as serotype of AAV virus injected [63] and copy number of ChR2 expressing vectors/alleles [131].

Another important issue that may induce different optogenetic effects is the difference in DRN 5-HT neural activity induced by optogenetic activation. In the previous study, ChR2(H134R) was expressed in SERT-Cre mice that received 25-Hz photostimulation. This induced strongly synchronized DRN 5-HT neural activity. Such

strong neural activity was mainly observed at the time of reward acquisition [99]. A recent study using the same transgenic mouse line showed that 20-Hz photostimulation of DRN 5-HT projections to the VTA induced dopamine release at the NAc and also had a reinforcing effect [191]. Several previous studies also showed that in the self-stimulation test, activation of DRN 5-HT neurons with 20-Hz photostimulation reinforced active nose poking to a port coupled with the stimulation [102, 131]. Photostimulation used in the previous study may affect active persistence and response vigor through the reward effect induced by dopamine release. On the other hand, in this study, we used ChR2(C128S), a step-type function opsin that induced less synchronized activity, i.e. approximately 6-Hz firing, and photostimulation to this ChR2 variant did not induce rewarding effects [128], suggesting that dopamine release is not induced by the stimulation. One possibility is that moderate and strong activation of DRN 5-HT neurons interacts with dopaminergic neurons differently, resulting in different effects on motivated behaviors. In order to examine, whether the optogenetic effect observed in the active nose-poking task is due to firing frequency or not, it is necessary to examine whether the same effect is found with our transgenic mice and stimulaion protocols in the active nose-poking task.

Another remarkable difference between our task and that of Lottem et al. [104] is the cost of abandoning a trial. The amount of travel cost to move to the next trial critically modulates animal decisions about whether to stay in the current trial/patch. Previous studies found that time and response in the current trial increased as the travel cost increased [114, 185]. In Lottem et al. [104], mice were required to travel a 30-cm passage to move to the next trial, which took approximately 3 s on average. On the other hand, in our task, the inter-trial intervals could function as a travel cost to start the next trial, which took at least 20 s. One possibility is that 5-HT neurons modulate active motor actions differently, depending on the magnitude of the travel cost. This possibility can be examined by testing optogenetic stimulation of 5-HT neurons in lever-pressing behaviors with inter-trial intervals of different lengths.

In conclusion, we showed that manipulation of DRN 5-HT activity has different effects on sustained motor action and stationary waiting for future rewards. Patience to wait and persistence to act for delayed reward acquisition are regulated differently by DRN 5-HT neurons. Advanced optical imaging from genetically tagged 5-HT neurons such as fiber photometry and endoscopic microscopy will enable us to identify different neural substrates in persistence to act and patience to wait.

Chapter 4

Control of model-based decision making by DRN 5-HT neurons

4.1 Introduction

In decision making, we select an action by estimating which action offers more rewards in the long run. In this process, there are mainly two distinct algorithms to decide: model-free and model-based RL systems [171]. Previous studies suggest that these two RL systems are implemented in the brain in parallel and that they cooperate to optimize policy [41, 48, 49]. Previous human studies suggest 5-HT neurons modulate arbitration between two valuation systems [201] However, this study reduced the 5-HT level systemically, and detailed neural substrates responsible for the arbitration were not examined. One recent study in mice found that optogenetic inhibition of DRN 5-HT neurons, but not MRN, disrupted model-based decision making in a reward devaluation task [137]. Although this study demonstrates the causal relationship between DRN 5-HT neurons and model-based decision making, it is difficult to examine which computation processes are disrupted by optogenetic inhibition: Does DRN 5-HT control both model-based and model-free decision making, or does it control only one of them? Another recent computational study proposed that DRN 5-HT neurons are also involved in model-based value estimation [125]. Behavioral studies examining how DRN 5-HT neurons control each computational process in model-based decision making are still few in number.

Here, to examine the computational role of DRN 5-HT neurons in model-based decision making, I used a mouse two-step decision-making task, a task developed to examine model-based decision making [3]. An important advantage of this task is that by combining it with RL models, descriptive models of decision-making processes, it is possible to quantitatively understand the computational processes of model-based decision making.

I trained mice to perform a two-step decision making task and tested the effects of photoinhibition of DRN 5-HT neurons on choice behavior and underlying computation of model-based decision making. I first found that mice use both model-free and model-based RL systems to make a choice. I also found that DRN 5-HT photoinhibition affected the features of model-based decision making. By fitting the behavioral data to

a model-free/model-based hybrid model, I found that photoinhibition of 5-HT neurons decreased the weight of model-based decision making.

4.2 Materials and Methods

4.2.1 Behavioral experiment

Animals. All experimental procedures were performed in accordance with guidelines established by the Okinawa Institute of Science and Technology Experimental Animal Committee. Six C57/BL background mice were used for baseline behavioral experiments. For the photoinhibition experiment, seven Tph2-ArchT-EGFP bi-transgenic mice were used. Three Tph2-tTA transgenic mice were used as control. All mice were housed individually at 24°C on a 12:12 h light: dark cycle (lights on 07;00-19;00 h). All behavioral training and photostimulation sessions were performed during the light cycle six days per week.

Behavioral apparatus. All training and photostimulation sessions were performed in custom-made operant boxes controlled by pyControl, a micropython-based behavioral control system [3]. On one side of the behavior boxes, five holes were located at the left, right, top, bottom, and center of the side. Each hole was equipped with an LED light. The upper and lower ports were connected to stainless tubes to deliver water rewards. A speaker to generate tone cues was located above the upper port.

Mouse two-step decision making task. I used the two-step decision making task developed in a previous study [3] with slight modification (Fig. 4.1). In the initial state, the center port lit up. Mice could start a trial with a nose poke in the center port. After the nose poke in the center port, the right and left ports lit up in the first step. After the first-step choice, either the upper or lower port illuminated, referred to as an up state or down state, respectively. The choice at the first step commonly led to one second-step state (up or down) with 80% and led to the other second-step state with 20%. Tones of different frequency signaled to which second-step state a trial would move. The state transition probability was fixed throughout training and photostimulation sessions. In the second step, mice were required to poke an illuminated port. After poking in the upper/lower port, mice could obtain water rewards with different probabilities. A water reward was delivered or not delivered with a 500-ms pure tone or white noise, respectively. After outcome delivery (reward or no reward), a 2-4-s inter-trial interval was inserted, and mice could start the subsequent trial after the center port lit up again. In the full task, mice were allowed to freely choose left or right in 75% of trials (referred to as free-choice trials) and forced to choose either port in the remaining 25% of trials (referred to as forced-choice trials). Reward probabilities in the up or down state were dynamically changed in block structures. In non-neutral probability blocks, the reward probability of one state was 80%, and that of the other state was 20%. In neutral probability blocks, reward probabilities of both states were 50%. In the non-neutral block, first-step choice behaviors were monitored as an exponential moving average of correct choices, which commonly led to a more rewarding state at the present non-neutral block. The non-neutral block was randomly changed to another block 5-15 trials after the moving average of correct choices surpassed 75%. The neutral block

was randomly switched to another block after 20-30 trials.

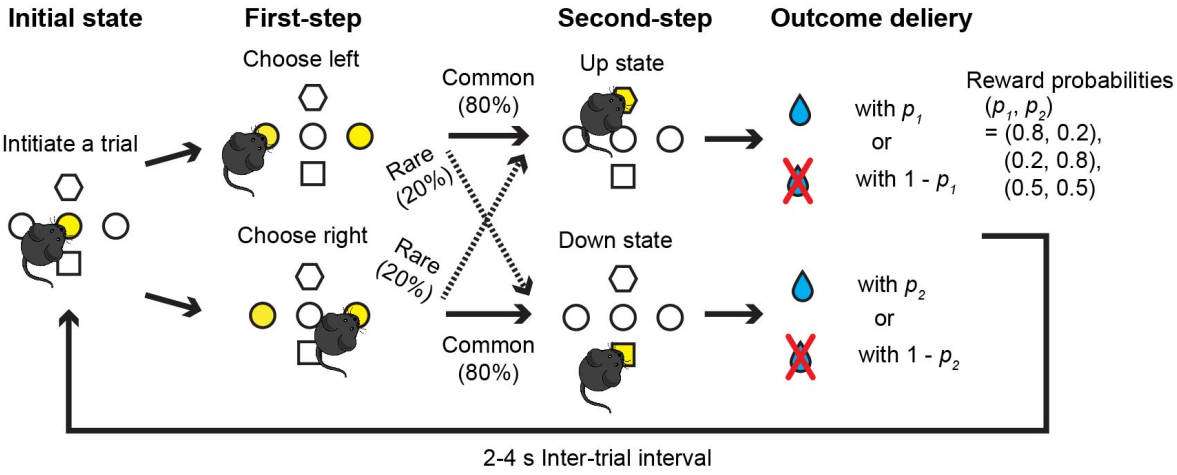

Figure 4.1: The schematic diagram of two-step decision making task.

Training for a two-step decision making task. Mice were restricted to free access to water 48 h before the first day of training. One day before the first training session, mice could freely access water. Optional ad libidum water access was given if needed to maintain their body weights at >85% of pre-restricted body weights. Training sessions consisted of 4 stages containing 11 substages in total. From Stage 1 to the middle of Stage 4, two 45-min sessions were performed per day, and a 90-min session was performed afterward. Each substage is designed as follows:

Stage 1: The purpose of Stage 1 was to establish an association between nose-poking in an upper or lower port and reward delivery. In Stage 1.1, either the upper or lower port was lit, and the mice deterministically obtained water rewards (15 µl) by a nose poke in the lit port. Once the mice could perform more than 50 trials in a session, training moved to Stage 1.2. In this stage, auditory cues were introduced to signal the reward delivery. The reward size was decreased to 12 or 10 µl. Fifty trials per session were required to move to the next stage.

Stage 2: In stage 2, probabilistic state transitions from the first step to the second step were inserted. At the first step, either the left or right port was lit. The state transition occurred by nose-poking in the lit port. At the second step, mice could deterministically obtain rewards by nose-poking in the lit port. After the mice could perform more than 50 trials, the training moved to stage 3.

Stage 3: In stage 3, the initial state was introduced. Mice were trained to nose-poke in the lit center port to initiate a trial. At the first step, either the left or right port was lit, and the trial moved to the second step by nose-poking in the lit port. After the state transition, by nose-poking in the lit port, mice could obtain rewards deterministically. The mice were required to perform at least 75 trials to move to stage 4.

Stage 4: Stage 4 consisted of 7 substages, and stage 4.7 was the full task used for photoinhibition sessions. In stage 4.1, all trials were forced-choice trials, and probabilistic reward delivery and reward block structure were introduced. In stages 4.2-4.6, by increasing the probability of free-choice trials and making reward probability more

Table 4.1: Training schedule of each substage at training stage 4

Stage	Forced-choice	Reward prob.	Block switch	Criteria to next stage
4.1	100%	Non-neutral 90% 70% Neutral 80% 80%	20-30 trials	$>$ 50 trials in a session
4.2	75%	Non-neutral 90% 50% Neutral 70% 70%	20-30 trials	$>$ 50 trials in a session
4.3	75%	Non-neutral 90% 30% Neutral 60% 60%	20-30 trials	$>$ 50 trials in a session
4.4	75%	Non-neutral 90% 10% Neutral 50% 50%	20-30 trials	$>$ 50 trials in a session
4.5	50%	Non-neutral 90% 10% Neutral 50% 50%	$>$ 75% correct choice	$>$ 50 trials in a session
4.6	25%	Non-neutral 90% 10% Neutral 50% 50%	$>$ 75% correct choice	5-6 blocks in a session
4.7 (Full)	25%	Non-neutral 80% 20% Neutral 50% 50%	$>$ 75% correct choice	

In the column of 'Reward prob.', each value indicates the reward probability at the two second-step states

stochastic, task complexity was increased. From stage 4.6, reward size was gradually decreased to 4 µl. Detailed information for each substage is described in Table 4.1. In stage 4.7, once the mice could experience more than 5 blocks per session, 2-3 days in a row, training was completed.

Surgical procedure for optic probe implantation. After training was completed, a craniotomy was performed to implant an optic probe (400 µm diameter, 0.48 NA, 5 mm length, Doric) above DRN. Mice were anesthetized with isoflurane (3% for induction and 1-1.5% during the surgery). Mice were placed on a stereotaxic stage, and their heads were fixed with ear bars. Then the skull was exposed with a blade, and a hole was made with a drill. Once the brain was exposed through the hole, the dura was removed using the tip of a needle, and an optic probe was lowered above the DRN through the hole (from bregma: posterior, -4.6mm; lateral, 0 mm; ventral, -2.6 mm). Light-sensitive adhesive and dental cement was applied to the skull to fix the implanted optic probe. Mice were returned to their home cages for recovery. I started to retrain the mice for behavioral tasks one week after the surgery and then proceeded to photostimulation sessions.

Photostimulation protocol. During photostimulation sessions, 470-mm blue and 590-mm yellow light stimulation was given, generated by an LED light source (Doric Lenses). Timing of stimulation was determined by TTL pulses controlled with pyControl. Either blue or yellow light stimulation was applied after outcome delivery until a first-step choice in the subsequent trial. Light intensity of blue and yellow light at the tip of the optical fibers was 1.2-1.5 mW and 2.8-3.2 mW, respectively. Yellow light stimulation was applied based on two conditions. 1) yellow light stimulation was applied only before free-choice trials. 2) After a free-choice trial preceded by yellow light stimulation, at least two subsequent trials were preceded by blue light stimulation. As a result, roughly 20% of trials were preceded by yellow light stimulation.

Immunohistochemistry. After the behavioral tests, mice were deeply anesthetized

with 100 mg/kg sodium pentobarbital i.p. and perfused with saline or PBS followed by 4% PFA/PB or 4% PFA/PBS. Their brains were removed immediately after perfusion and immersed in fixative solution overnight. Then, 50 μm coronal slices were cut using a vibratome (VT1000S, Leica), and the implantation site of optic probes was confirmed according to the mouse brain atlas [64].

In order to confirm ArchT expression after the task, brain slices were incubated with primary antibodies for two nights. Slices were then rinsed with PBS and incubated with secondary antibodies for two nights. After incubation and rinsing, slices were mounted onto slide glasses. Fluorescent images shown in Figure 4.4a were acquired using spinning disc confocal microscopy (SD-OSR, Olympus). As primary antibodies, I used anti-Tph (1:250, sheep polyclonal, Merck Millipore, AB1541) and anti-GFP (1:500, chicken polyclonal, Abcam, ab13970) as a marker for 5-HT neurons and ChR2-EGFP or ArchT-EGFP neurons, respectively. For secondary antibodies, anti-sheep and anti-chicken antibodies, respectively conjugated with Alexa flour 594 and 488 were used. Antibodies were diluted in staining buffer containing 10 mM HEPES, 20 mM NaCl, and 10% Triton X-100. The pH of the staining buffer was adjusted to 7.4 in advance.

4.2.2 RL model analysis

Description of RL models. I considered four RL models to decide the first-step choices: 1) model-free RL strategy, 2) model-based RL strategy, 3) model-free/model-based hybrid strategy, and 4) an inference strategy. I denote the first-step choice in the t th trial as $c_t \in [L, R]$, the second-step state as $s_t \in [U, D]$, and the reward/reward omission as $r_t \in [0, 1]$. In model-free RL strategy, Action values of the first-step actions taken at t th trial $Q_{mf,t}(c_t)$ are updated as follows:

$$Q_{mf,t}(c_t) \leftarrow (1-\alpha)Q_{mf,t-1}(c_t) + \alpha(\lambda r_t + (1-\lambda)V_{t-1}(s_t))$$

α and λ denote the learning rate and eligibility trace, respectively. $V_{t-1}(s_t)$ denotes the state value of the second-step state reached at t-1 th trial. The value of the second-step state reached was updated after outcome delivery as follows:

$$V_t(s_t) \leftarrow (1-\alpha)V_{t-1}(s_t) + \alpha r_t$$

On the other hand, the value of the first-step action unchosen $Q_{mf,t}(c'_t)$ and the value of the second-step state unattained $V_t(s'_t)$ decayed toward zero with a forgetting rate f as follows:

$$Q_{mf,t}(c'_t) \leftarrow (1-f)Q_{mf,t-1}(c'_t)$$
$$V_t(s'_t) \leftarrow (1-f)V_{t-1}(s'_t)$$

In model-based RL strategy, values of the second-step states were updated as in the model-free RL strategy. Action values of first-step actions after t th trial was calculated as the sum of the state values of the second-step state weighted by the state transition probabilities. For example, the action value of the left choice was calculated as follows:

$$Q_{mb,t}(c = L) \leftarrow p(s = U|c = L)V_t(s = U) + p(s = D|c = L)V_t(s = D)$$

$p(s|c)$ denotes the state transition probabilities to each second-step state after the first-step actions. I used $p(s|c) = 0.8$ for common transition and $p(s|c) = 0.2$ for rate transition. The action value of the right choice was also computed in the same manner. Based on these action values, net action values $Q_{net,t}$ are calculated as follows:

$$Q_{net,t}(c = L) \leftarrow \beta_{mf} Q_{mf,t}(c = L) + \beta_{mb} Q_{mb,t}(c = L) + P\bar{c} + B$$

$$Q_{net,t}(c = R) \leftarrow \beta_{mf} Q_{mf,t}(c = R) + \beta_{mb} Q_{mb,t}(c = R)$$

β_{mf} and β_{mb} control how much the agents rely on each strategy. Therefore, the model-free RL strategy uses $\beta_{mb} = 0$ and the model-based RL strategy uses $\beta_{mf} = 0$. In a model-free/model-based hybrid RL strategy, both β_{mf} and β_{mb} are positive values. In the net action value of the left choice, two additional terms are added in order to model the animals' tendency to repeat the same choice as in the previous trial (i.e., perseveration) and the preference for a specific choice (i.e., bias). In detail, perseveration is modeled as follows:

$$\bar{c} = \begin{cases} 0.5 & \text{for } c_t = L \\ -0.5 & \text{for } c_t = R \end{cases}$$

P and B denote the strength of perseveration and bias, respectively. If B is positive/negative, agents have a preference for left/right choice, regardless of the action values of each choice. Based on the net action value, the probability of each action π is calculated as follows:

$$\pi(c_{t+1} = L) = \frac{\exp(Q_{net,t}(c = L))}{\exp(Q_{net,t}(c = L)) + \exp(Q_{net,t}(c = R))}$$

$$\pi(c_{t+1} = R) = \frac{\exp(Q_{net,t}(c = R))}{\exp(Q_{net,t}(c = L)) + \exp(Q_{net,t}(c = R))}$$

The inference strategy model can behave similarly to model-based strategies even without the internal model of transition probability [2]. In a recent study using the same configuration of mice, a two-step task evaluated various models based on the inference strategy and showed that one of them fit better than the model-free/model-based hybrid strategy [17]. Here, I evaluated the best fit model in the previous study. This model has two components: inference and reward-as-cue. These two components estimate action values in different ways. In the inference component, when agents obtain rewards, they learn the reward probability of second-step states as follows:

$$I_t \leftarrow (1 - \alpha_r) I_{t-1} + \alpha_r S_t$$

I denotes a subjects' belief about the reward probability of the up state and directly determined the value of the up state. α_r denotes the learning rate to update the belief. S indicates the second-step state reached at t th trial, taking 1 at the up state and 0 at the down state. Because the task has anti-correlated reward probabilities (i.e., if the up state has a higher probability, the down state has a lower reward probability),

the value of each second-step state was determined as follows:

$$V_{inf,t}(U) = I_t$$

$$V_{inf,t}(D) = 1 - I_t$$

Second-step state values determine first-step action values, based on which first-step actions commonly lead to specific second-step states. For example, if the left choice commonly leads to the up state and the right choice leads to the down state, first-step action values for the subsequent trial were determined as follows:

$$Q_{inf,t}(L) = I_t$$

$$Q_{inf,t}(R) = 1 - I_t$$

In the reward-as-cue component, when the agents obtain a reward at a second-step state, they reinforce the first-step action, which commonly leads to the second-step state reached. For example, if the task has a structure such that the left choice commonly leads to the up state and the right choice to the down state, and the agents obtain the reward in the up state, the first-step action values at the subsequent next trial are updated as follows:

$$Q_{rc,t}(L) = 1$$

$$Q_{rc,t}(R) = 0$$

Based on the action values from the inference and reward-as-cue component, the net action values for first-step actions are calculated as the weighted sum of action values from these two components with perseveration and bias.

$$Q_{net,t}(c = L) \leftarrow \beta_{inf} Q_{inf,t}(c = L) + \beta_{rc} Q_{rc,t}(c = L) + P\bar{c} + B$$

$$Q_{net,t}(c = R) \leftarrow \beta_{inf} Q_{inf,t}(c = R) + \beta_{rc} Q_{rc,t}(c = R)$$

The probability of each choice in the subsequent trial is calculated based on the softmax decision rule mentioned above.

Model evaluation and parameter estimate. I evaluated each model based on how accurately the model predicts choice behaviors, as in previous studies [87, 119]. To quantify the accuracy of the prediction, I calculated the normalized likelihood of observed first-step choices and evaluated the models using leave-one-out cross-validation. For a single t th trial, the likelihood $Z(t)$ was calculated as the probability of the first-step choice taken at t th trial given the choice history till t - 1 trial. The normalized log-likelihood of one session was calculated as follows:

$$\frac{1}{T_s} \sum_{t=1}^{T_s} \log Z(t)$$

T_s denotes the number of free-choice trials in one session.

For model evaluation, when I had behavioral data of n sessions from the same mouse, n - 1 sessions were assigned as training data used for the parameter estimate of

the model. The remaining session was used as test data for calculation of normalized log-likelihood. This process was repeated n times so that all session data were assigned as test data. Parameter values that maximize a posteriori (MAP) were estimated. The MAP estimate was performed using an optimizing function in a python library called pyStan. The prior distribution of each parameter was set as follows:

$$\alpha, f, \lambda, \alpha_r \sim beta(a = 2, b = 2)$$

$$\beta_{mf}, \beta_{mb}, \beta_{inf}, \beta_{rc} \sim gamma(\alpha = 2, \beta = 3)$$

$$P, B \sim student_t(\nu = 4, \mu = 0, \sigma = 2.5)$$

In the parameter estimate of the photoinhibition testing sessions, I assumed that parameter values under yellow and blue stimulation were different. Therefore, the number of free parameters was doubled compared to models for baseline data.

Simulation of behavioral sessions. To confirm whether the selected model behaves similarly to actual mice, I simulated behavioral sessions of baseline experiments. After the parameter estimate of each baseline dataset, behavioral sessions were simulated. Each session had the same length of trials as actual data. I also performed simulation of photoinhibition sessions using parameter estimate of each ArchT mice dataset. For simulation of photoinhibition session, Each session had 1000 trials. In addition, assuming that DRN 5-HT photoinhibition affected the weight on model-based decision making, I simulated photoinhibition sessions. The simulated sessions include no-inhibition and inhibition trials, the same as the actual photoinhibition sessions. For the no-inhibition trials, parameter values were set to the median of MAP estimate values for each mouse subject of the baseline experiment. For inhibition trials, the value of β_{mb} was set to zero. Similarly, photoinhibition sessions were simulated with the forgetting rate set to 0.01 or 0.99. For each mouse subject, 500 sessions, each of which has 450 trials, were simulated, and stay probabilities after the inhibition and no-inhibition trials were analyzed.

4.3 Results

4.3.1 Mice used both model-free and model-based decision making in the two-step task

The first-step choice behaviors and stay probability: To confirm how mice perform the task, I first trained six mice without implanting optic probes. Detailed information about baseline session data used for analysis is listed in Table 4.2. The moving average of making left choices changed adaptively based on which first step choice commonly led to more rewarding second-step state (Fig. 4.2a). To characterize their behaviors, I measured stay probability, how frequently mice repeat the same first-step choice as in the preceding trial. When mice experienced rewarded and common-transition trials, they were likely to repeat the same first-step choice (Fig. 4.2b). On the other hand, the stay probability was relatively low when they experienced reward and rare-transition trials. This result suggests that mice seemed to choose the subsequent action dependent on transition type, which is characteristic of model-based decision making.

Table 4.2: Summary of the baseline session data.

# mice	# sessions	# free-choice trials	# forced-choice trials	# blocks/session	# trials/block
6	89	26222	8735	10.5	35.2

Evaluation of RL models: To understand the computational process of how mice choose actions, first-step choice behaviors were fit to RL models. I examine four decision making strategy models: model-free RL strategy, model-based RL strategy, model-free/model-based hybrid strategy, and inference strategy (see section 4.2.1 for the model description). Behavioral data were divided into training and test data. Using the training data, values of free parameters were estimated using MAP estimate. The normalized log-likelihood of choice behaviors in the test data was calculated with the estimated values. The higher likelihood indicated the model could predict mouse choice behaviors of the test data more accurately. Data from all but one session were assigned as training data and data from the remaining session as test data. I repeated this process so that all sessions were assigned as test data (i.e., leave-one-out cross-validation). The normalized log-likelihood computed from each test data was averaged, and the average was exponentiated. This value indicates the average performance of the model and was computed for each model for each mouse. The model-free/model-based hybrid strategy model had the largest model accuracy in 4 of 6 mice (Hyb in Fig. 4.2c), and in the remaining mouse, the model accuracy of the inference strategy (INF+RC in Fig. 4.2c) slightly outweighed that of the hybrid strategy. On average, the hybrid strategy model had the largest likelihood in the prediction of choice behaviors (Fig. 4.2c). Some studies showed improved model fitting by addition of the value-forgetting process [3, 175, 176]. To examine this, I also compared the normalized likelihood between hybrid strategies with or without the forgetting process. Removal of forgetting decreased the model accuracy (Hyb + no forget in Fig. 4.2c).

To verify whether the selected model behaves similarly to actual mice, I simulated baseline sessions using the hybrid strategy model and calculated stay probabilities. For each session, free parameter values were estimated, and baseline sessions were simulated using the estimated value and the same number of trials. The mean of parameter values across subjects is listed in Table 4.3. Simulated stay probability looked similar to the actual behavioral data (Fig. 4.2b for actual behavior and Fig. 4.2d for simulated behavior). These results suggest that mice were likely to perform two-step tasks using both model-free and model-based decision making.

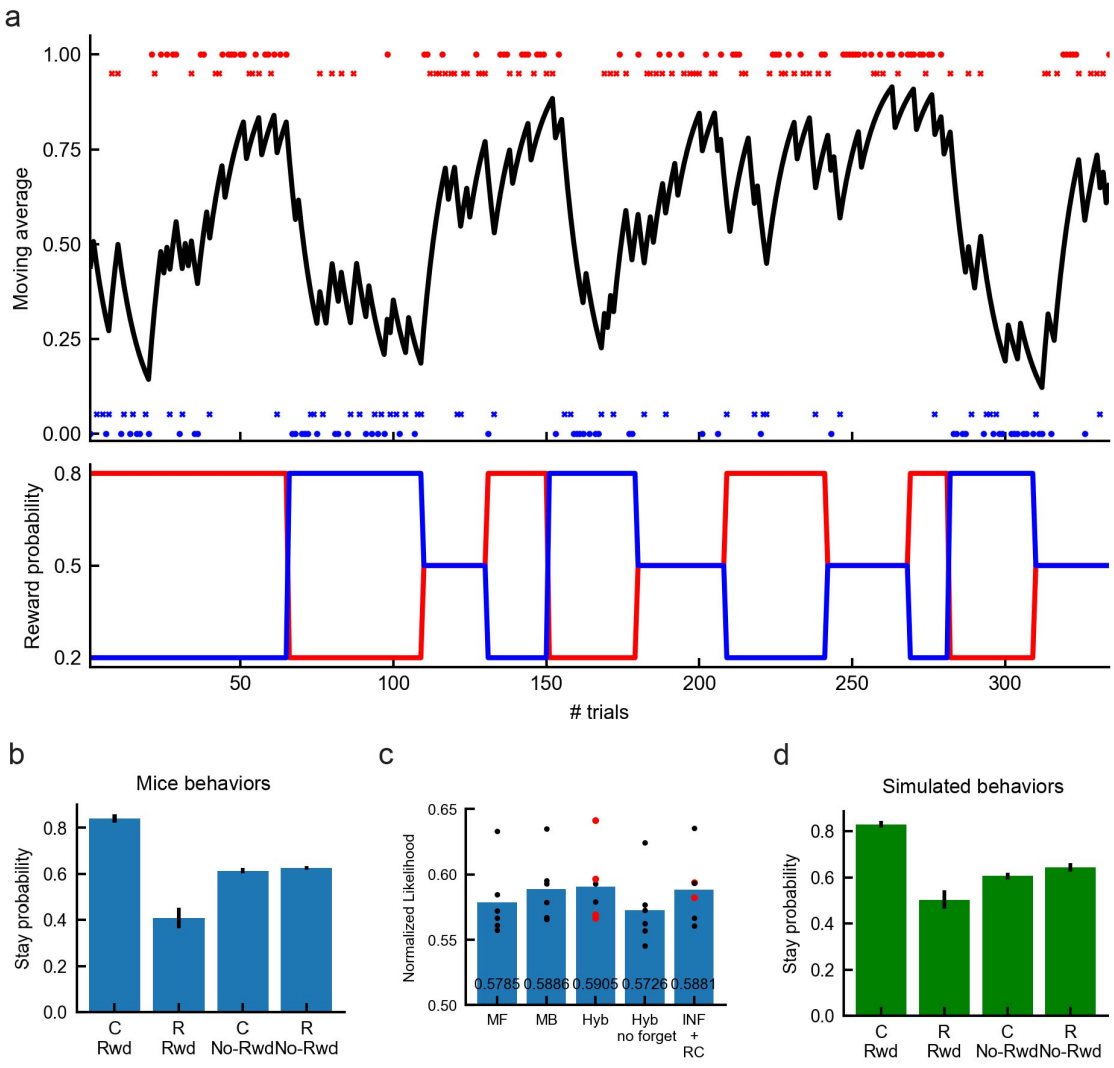

Figure 4.2: Behaviors in baseline sessions. a: The representative session. In the top panel, the history of first-step choices is shown. Red/blue dots indicate rewarded free-choice trials after choosing left/right, and red/blue crosses indicate non-rewarded free-choice trials after left/right choice. The black line indicates the exponential moving average of the left choice. In the bottom figure, red/blue lines indicate the reward probabilities of the second-step states commonly led by left/right choice. b: Stay probabilities after the four combinations of outcome and transition in a previous trial ($n = 6$ mice). c: The accuracy of each model in each mouse ($n = 6$ mice). Black dots indicate the averaged likelihood for each mouse, and red dots indicate the highest likelihood of each mouse. The height of blue bars and the values inside the bars indicate the mean across mice. d: Stay probabilities in sessions simulated by the model-free/mode-based hybrid model. Error bars represent SEM in all the graphs.

Table 4.3: MAP estimate values in the hybrid model in baseline sessions.

Parameter	Mean ± s.d
α	0.336 ± 0.111
f	0.448 ± 0.0653
λ	0.481 ± 0.0759
β_{mf}	1.688 ± 0.580
β_{mb}	3.645 ± 0.770
P	0.482 ± 0.251
B	0.0944 ± 0.129

4.3.2 Predicted effect of disrupted model-based decision making on choice behaviors

Previous studies have proposed that 5-HT controls the weight of model-based decision making to guide actions [137, 201]. Therefore, I predicted that DRN 5-HT photoinhibition would decrease the weight of model-based decision making. To examine how the expected effect of photoinhibition changes choice behaviors in the task, I simulated photoinhibition sessions with β_{mb} set to 0 in inhibition trials. Disrupted model-based decision making significantly changed stay probabilities except after common non-rewarded trials (Fig. 4.3a). Remarkably, the stay probability after rare rewarded trials was significantly increased in the simulated photoinhibition trials, and choice behaviors became independent of the type of transitions. In addition, other studies showed that 5-HT regulates value learning processes in dynamic environments [69, 85]. Therefore, another possible effect of optogenetic inhibition is that DRN 5-HT inhibition affects value learning and/or forgetting processes and changes choice behaviors. To examine the effect of changed value update processes, I simulated photoinhibition sessions with the learning rate (α) or forgetting rate (f) set to very small or large values (Fig. 4.3b for learning rate and c for forgetting rate). A smaller learning or forgetting rate significantly increased the stay probability, similarly to disrupted model-based decision making, although the effect seems much smaller.

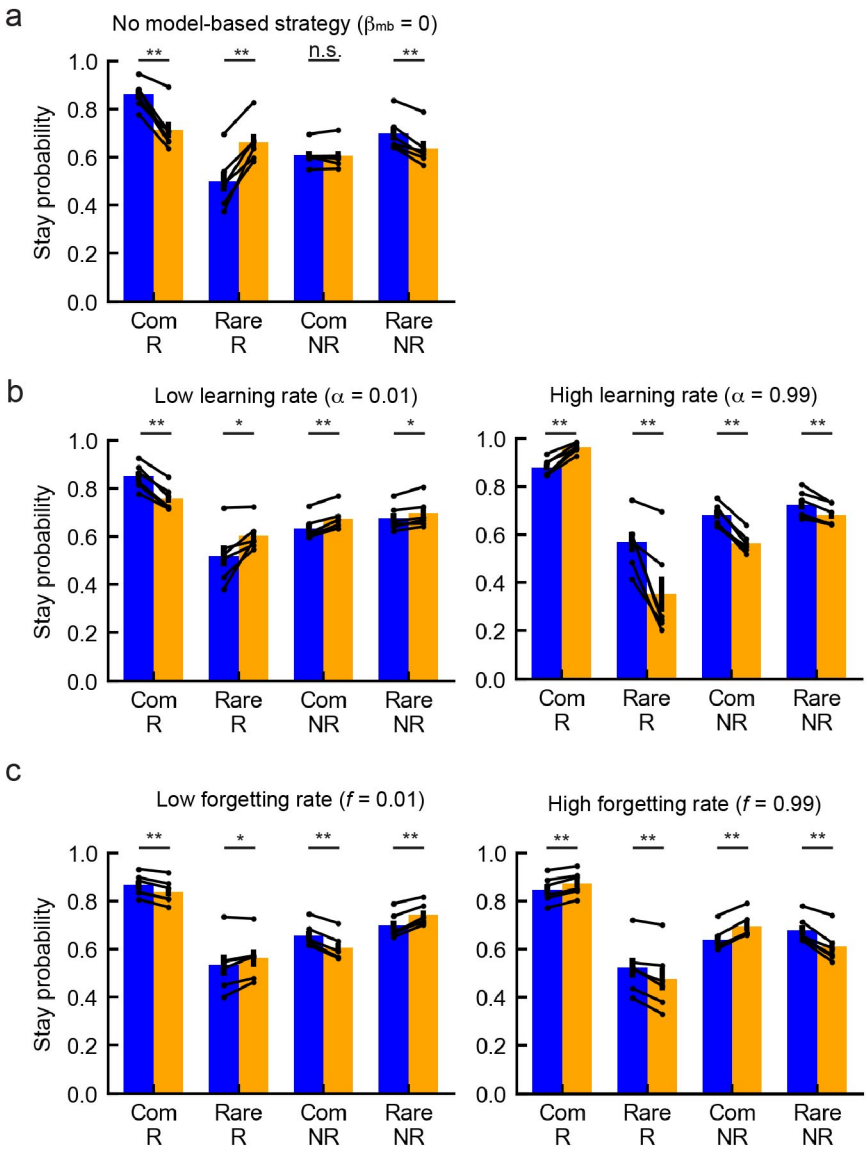

Figure 4.3: Stay probabilities in simulated photoinhibition testing sessions. a: β_{mb} set to 0 in inhibition trials. Black lines indicate stay probabilities from each simulated mouse subject. Blue and yellow bars indicate the mean across the simulated subjects in no-inhibition and inhibition trials, respectively. ** indicates $p < 0.01$. n.s. indicates no significance ($p > 0.05$) by paired t-test. b: α set to very small (0.01) or large (0.99) values in inhibition trials. Black lines indicate stay probabilities from each simulated mouse subject. Blue and yellow bars indicate the mean across the simulated subjects in no-inhibition and inhibition trials, respectively. ** indicates $p < 0.01$. n.s. indicates no significance ($p > 0.05$) by paired t-test. c: f set to very small (0.01) or large (0.99) values in inhibition trials. Black lines indicate stay probabilities from each simulated mouse subject. Blue and yellow bars indicate the mean across the simulated subjects in no-inhibition and inhibition trials, respectively. ** indicates $p < 0.01$. n.s. indicates no significance ($p > 0.05$) by paired t-test. Error bars represent the SEM.

4.3.3 Photoinhibition and histological confirmation of optic probes

To examine how DRN 5-HT neurons are causally involved in model-based decision making, DRN 5-HT neural activities were silenced using photostimulation in Tph2-ArchT bi-transgenic mice during the two-step task. I trained 7 Tph2-ArchT (referred to as ArchT mice) and 3 Tph2-tTA mice (referred to as control mice) to perform the two-step task and then implanted optic probes above their DRNs (Fig. 4.4a, see section 4.2.1 for the detail of histological experiment procedure). During testing sessions, either blue- or yellow-light stimulation was continuously applied from the outcome delivery of a trial to the first-step choice in the next trial to inhibit or not inhibit DRN 5-HT neural activities (Fig. 4.4b, see section 4.2.1 for details of the photoinhibition protocol).

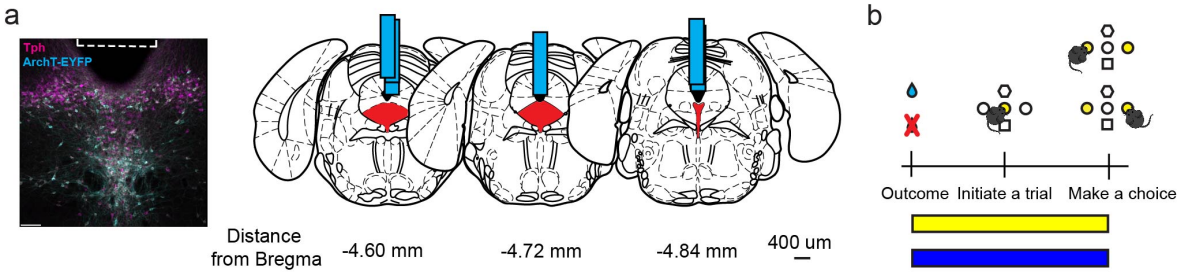

Figure 4.4: The implantation site of optic probes in ArchT mice and photoinhibition protocol. a: The left image shows representative ArchT expression in ArchT mice. The right panel shows the implantation site of optic probes on coronal views of the mouse brain (adapted from [64]). Blue rectangles indicate the tracks of the implanted optic probes. Red filled areas indicate DRN. b: Timing of optic stimulation in a trial.

4.3.4 The effect of photoinhibition of DRN 5-HT neurons on first-step choice behaviors

Photoinhibition testing sessions: One session was performed per day for three weeks. In total, 16-18 sessions were performed with each mouse. ArchT expression was confirmed after repeated long-term stimulation, suggesting that the stimulation protocol did not cause discernible damage to 5-HT neurons. 122 behavioral sessions from ArchT mice and 53 sessions from control mice were used for analysis. Detailed information of photoinhibition testing sessions is given in Table 4.4.

Table 4.4: Summary of the photoinhibition testing session data.

	# mice	# sessions	# free-choice trials	# forced-choice trials	# blocks/session	# trials/block
ArchT	7	122	blue 36558 yellow 5724	blue 9135 yellow 4959	13.7	32.2
Control	3	53	blue 16052 yellow 2436	blue 3915 yellow 2244	14.2	31.2

In the column of # free-choice trials and # forced-choice trials, the number of each type of trial followed by blue/yellow light stimulation is listed.

Stay probability analysis: To capture characteristics of choice behaviors, I calculated the stay probabilities. Because a previous study showed that acute activation and chronic activation (15-min activation for three weeks) of DRN 5-HT neurons displayed the opposite behavioral effect [32], to examine whether the photostimulation effect is time-dependent, stay probability for each testing week was calculated. Stay probability after rare rewarded trials consistently decreased throughout testing weeks in the present results (Fig. 4.5 ArchT mice). This tendency was also found in the simulation above, after the disruption model-based decision making or decrease in learning or forgetting rates. Therefore, these results indicate that DRN 5-HT photoinhibition may reduce the feature of model-based decision making. However, in an analysis of stay probability after rare rewarded trials with two-way repeated measures ANOVA, there were no significant main effects of light (two levels within-subject factors; yellow and blue, $F(1,6)=3.1$, $P=0.13$) and testing week (three levels within-subject factors; week 1, week 2, week 3, $F(2,12)=1.2$, $P=0.34$).

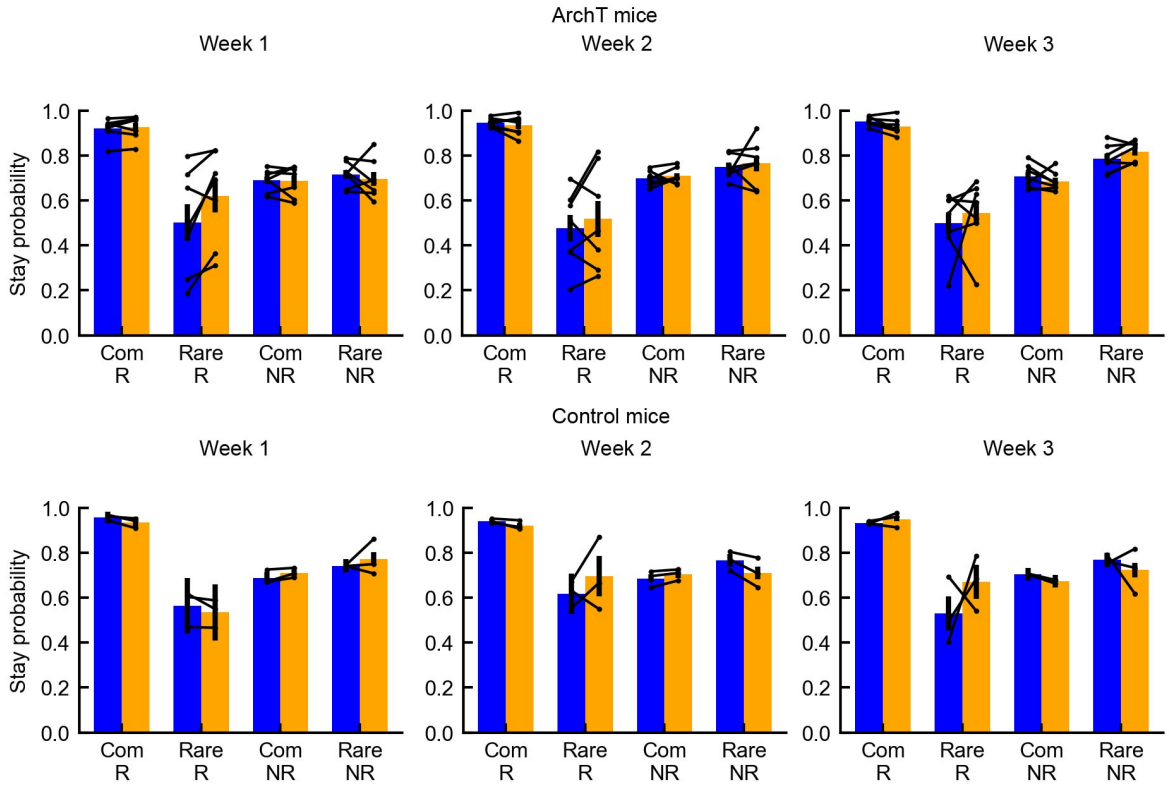

Figure 4.5: DRN 5-HT neurons reduced the feature of model-based decision making in stay probabilities. Stay probabilities of ArchT and control mice (n = 7 ArchT and 3 control mice) in each testing week. Heights of blue and orange bars indicate the means under blue- and yellow-light stimulation. Error bars represent the SEM in all graphs.

4.3.5 The effect of DRN 5-HT photoinhibition on the decision-making process in the two-step task

Evaluation of RL models in photoinhibition sessions

To quantitatively examine the effect of photoinhibition on the computational process of decision making, first-step choice behaviors were analyzed using RL models. RL Models were evaluated by week of testing sessions. In all testing weeks, the hybrid model had the highest mean accuracy in predicting choice behaviors (Fig. 4.6). Therefore, the model-free/model-based hybrid model was fitted to behavioral data in later sections.

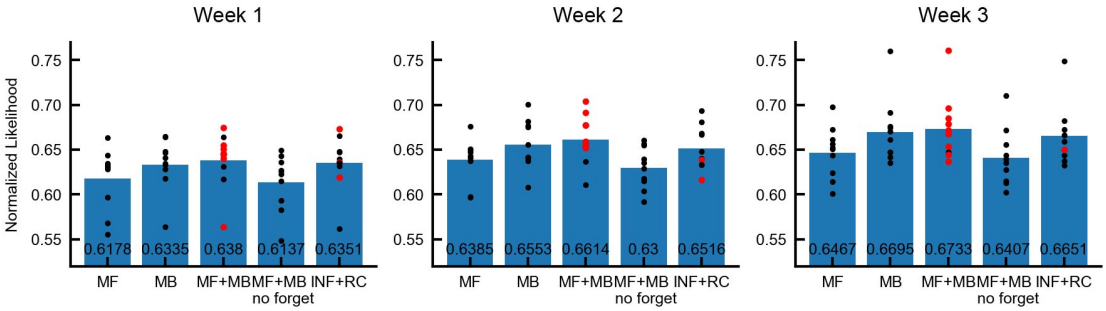

Figure 4.6: Accuracy of RL models in photoinhibition testing sessions. Model accuracy during each testing week of photoinhibition sessions (n = 7 ArchT and 3 control mice). Heights of blue bars and values inside the bars indicate means among mice. The red dots indicate the highest likelihood of each mouse.

Fitting of model-free/model-based hybrid model to photoinhibition sessions

The hybrid model was fitted session by session, and parameter values were estimated by MAP estimate. The median across sessions was calculated for each week at each mouse. As in the predicted effect of photoinhibition, here, I describe the effect of photoinhibition on the weight of model-based decision making, the balance between model-free and model-based strategy, learning rate, and forgetting rate. Other parameters did not show a remarkable photoinhibition effect.

The weight of model-based strategies: To examine whether DRN 5-HT neurons control the use of model-based decision making, I analyzed the photoinhibition effect on β_{mb}, the weight of model-based decision making (Fig. 4.7a). In ArchT, β_{mb} was consistently reduced throughout testing weeks. In analysis of ArchT mice data with two-way repeated measures ANOVA, there were significant main effects of light (two levels within-subject factors; yellow and blue, $F(1,6) = 10.1$, $P = 0.019$) and testing week (three levels within-subject factors; week 1, week 2, week 3, $F(2,12) = 18.5$, $P = 0.00022$). However, there was no significant main effect of interaction (light x week, $F(2,12) = 1.32$, $P = 0.30$). On the other hand, in control mice, there was no significant main effect on light ($F(1,2) = 0.45$, $P = 0.57$).

The balance between model-free and model-based decision making: To examine whether photoinhibition affected the balance between model-free and model-based decision making, I calculated the angle between data points and β_{mf} axis in the space

of weight parameters(Definition given in the right graph in Fig. 4.7a). An angle larger than 45° indicates greater reliance on model-based over model-free decision making and vice versa. In ArchT, angle was consistently decreased in photoinhibition trials throughout testing weeks (Fig. 4.7b). In the analysis of ArchT mice data with two-way repeated measures ANOVA, there was significant main effect of light (two levels within-subject factors; yellow and blue, $F(1,6) = 8.02$, $P = 0.030$), but no significant main effect of testing week (three levels within-subject factors; week 1, week 2, week 3, $F(2,12) = 1.91$, $P = 0.19$). There was no significant main effect of interaction (light x week, $F(2,12) = 17.8$, $P = 0.66$). On the other hand, in control mice, there was no significant main effect on light ($F(1,2) = 1.76$, $P = 0.32$). Analysis of the weight of model-based decision making suggests that photoinhibition of DRN 5-HT neurons reduced the weight on model-based decision making and shifted the balance toward model-free decision making.

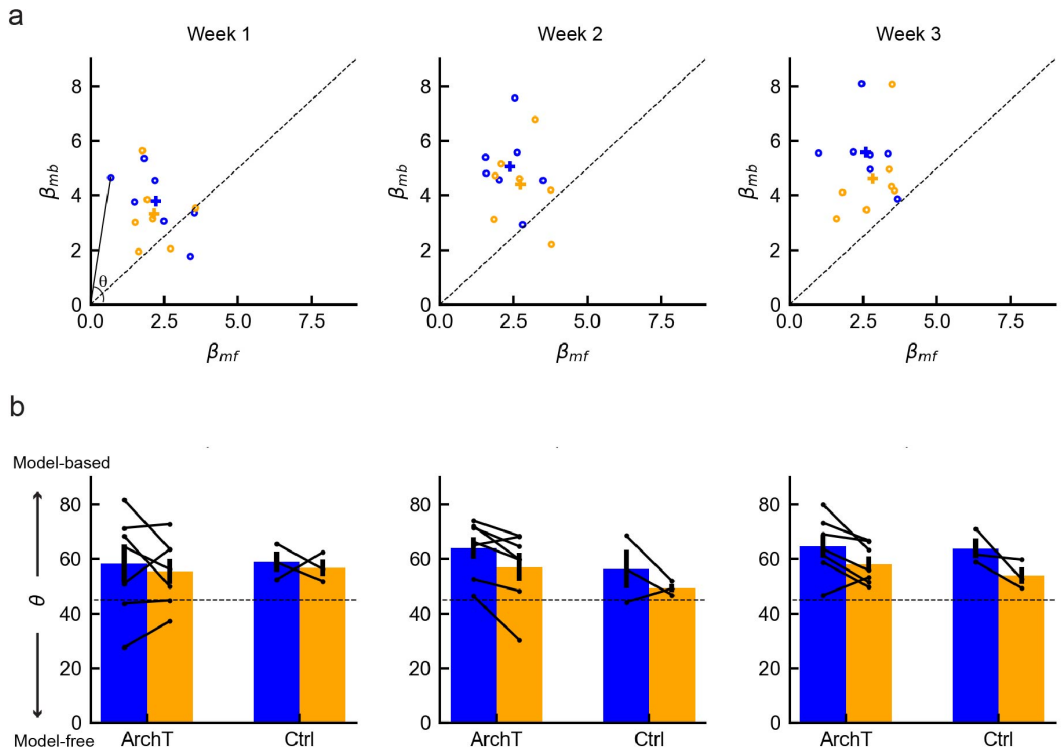

Figure 4.7: The effect of DRN 5-HT photoinhibition on the weight of model-based decision making. a: The weights of model-free (β_{mf}) and model-based (β_{mb}) decision making in ArchT mice ($n = 7$ ArchT mice) at each testing week. Blue and yellow open circles indicate data of individual ArchT mice. Blue and orange crosses indicate the mean across subjects. b: The balance between model-free and model-based decision making in ArchT and control mice ($n = 7$ ArchT and 3 control mice). The balance of the two systems is represented by the angle between line to the data and β_{mf} axis in the weight space in Fig. 4.7a. Heights of blue and orange bars indicate the mean across mice in blue and yellow light stimulation, respectively. Black dots indicate the estimated value of individual subjects. Dashed lines indicate 45° at which model-free and model-based decision making are equally used to guide actions. Error bars represent the SEM in all bar graphs.

Learning rate: The simulation suggests that the change in value update processes also affects choice behaviors. Therefore, I first examined the effect of photoinhibition on learning rate using two-way repeated measures ANOVA (Fig. 4.8a). In ArchT mice, there were no significant main effects of light (two levels within-subject factors; yellow and blue, $F(1,6) = 2.68$, $P = 0.15$) and testing week (three levels within-subject factors; week 1, week 2, and week 3, $F(2,12) = 2.33$, $P = 0.14$). In control mice, although there was a significant main effect of testing week ($F(2,4) = 20.13$, $P = 0.0082$), there was no significant main effect of light ($F(1,2) = 0.341$, $P = 0.618$) and no significant effect of interaction (light x week, $F(2,4) = 0.579$, $P = 0.60$). These results suggest that photoinhibition did not affect learning rate.

Forgetting rate: Lastly, I analyzed the photoinhibition effect on forgetting rate with two-way repeated measures ANOVA (Fig. 4.8b). In ArchT mice, there were no significant main effects of light (two levels within-subject factors; yellow and blue, $F(1,6) = 0.052$, $P = 0.83$) and testing week (three levels within-subject factors; week 1, week 2, and week 3, $F(2,12) = 2.92$, $P = 0.093$). There was also no effect on interaction (light x week, $F(2,4) = 0.85$, $P = 0.49$). In control mice, although there was a significant main effect of testing week ($F(2,4) = 8.28$, $P = 0.038$), there was no significant main effect of light ($F(1,2) = 1.35$, $P = 0.37$) and no significant effect of interaction (light x week, $F(2,4) = 0.85$, $P = 0.49$). These results suggest that photoinhibition did not affect forgetting rate.

Taken together, RL model analysis shows that DRN 5-HT inhibition reduced reliance on model-based decision making to guide actions.

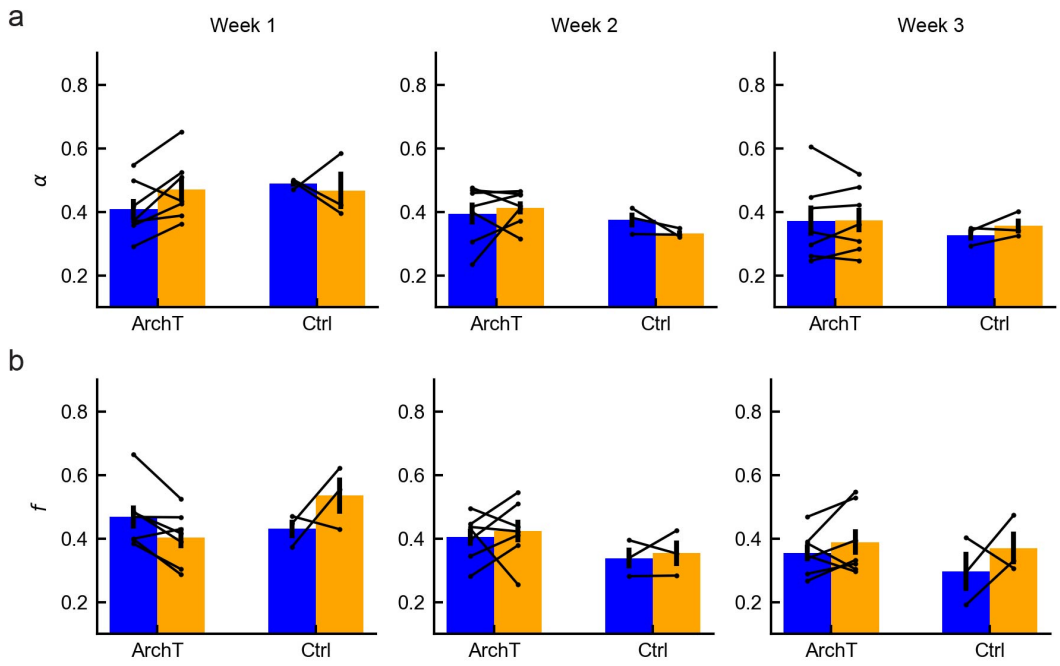

Figure 4.8: The effect of DRN 5-HT photoinhibition on value learning/forgetting. a: Estimated value of the learning rate in ArchT and control mice ($n = 7$ ArchT and 3 control mice) at each testing week. Heights of blue and orange bars indicate the mean across mice under blue-light and yellow-light stimulation, respectively. b: Estimated value of the forgetting rate in ArchT and control mice ($n = 7$ ArchT and 3 control mice) at each testing week. Heights of blue and orange bars indicate the mean across mice in blue- and yellow-light stimulation, respectively. Error bars represent the SEM in all graphs.

Simulation of photoinhibition sessions using estimated parameter values

In order to confirm whether parameter values estimated from behavioral sessions data can reproduce choice behaviors similar to actual behaviors, I simulated photoinhibition sessions using estimated parameter values. Overall, simulated stay probabilities look similar to actual choice behaviors. Stay probability after rare rewarded trials largely tended to decrease in photoinhibition trials, although the change in simulated behaviors was not as obvious as actual behaviors.

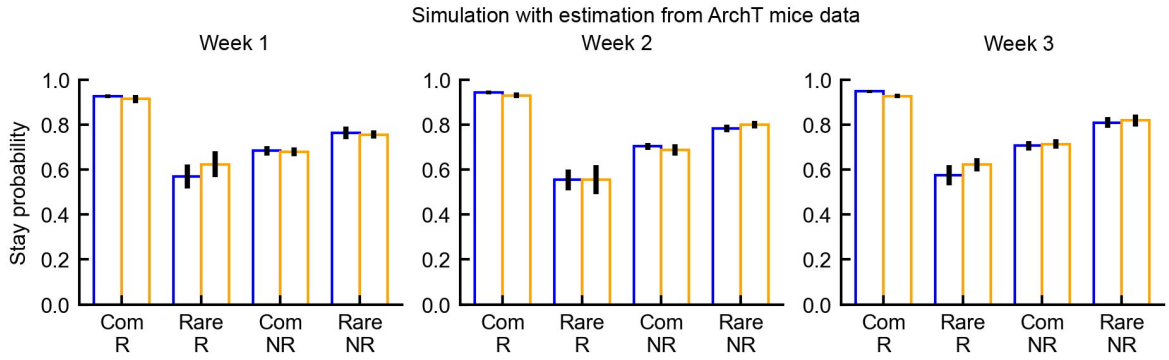

Figure 4.9: Simulation of photoinhibition sessions using estimated parameter values. Stay probabilities of simulated sessions using estimated parameters from ArchT mice in each testing week. Heights of blue and orange bars indicate the means under blue- and yellow-light stimulation. Error bars represent the SEM in all graphs.

4.3.6 The effect of photoinhibition of DRN 5-HT neurons on response time

To further examine the effect of photoinhibition on behavior, I measured response time at different states in a trial: latency to initiate a free-choice trial and the time to make a first-step choice.

The latency to initiate a free-choice trial: The speed of action is latent variable for decision making and could reflect the value of the current state [104, 190]. In order to examine how much mice were motivated to start a trial and how this motivation is affected by photoinhibition, the duration from entry to the initial state (illumination of the center port) to a nose poke in the center port was measured. Yellow-light stimulation increased the latency to initiate a trial in ArchT mice across weeks (Fig. 4.10a). I analyzed the response time using two-way repeated measures ANOVA. In ArchT mice, there were significant main effects of light (two levels within-subject factors; yellow and blue, $F(1,6) = 29.29$, $P = 0.0016$) but no significant effect on time (three levels within-subject factors; week 1, week 2, and week 3, $F(2,12) = 6.10$, $P = 0.015$) and interaction (light x week, $F(2,12) = 1.08$, $P = 0.37$). In control mice, there were no significant main effects on light ($F(1,2) = 11.18$, $P = 0.079$) and time ($F(2,4) = 6.39$, $P = 0.057$), and the effect of interaction was not significant neither ($F(2,4) = 0.46$, $P = 0.66$).

The time to make a first-step choice: One important feature of model-based decision making is that model-based planning requires mental simulation to make decisions

[151]. Therefore, I predicted that if inhibition of DRN 5-HT disrupted model-based decision making, the inhibition reduced deliberative process and reduced the time to make decisions. To examine this, I measured the time to make a first-step choice, the duration from trial initiation by nose poke to the center port to a first-step choice by a nose poke in left or right port. Yellow-light stimulation decreased the latency to initiate a trial in ArchT mice across weeks (Fig. 4.10b). I analyzed the response time using two-way repeated measures ANOVA. In ArchT mice, there were significant main effects on light (two levels within-subject factors; yellow and blue, $F(1,6) = 23.47$, $P = 0.0029$) and time (three levels within-subject factors; week 1, week 2, and week 3, $F(2,12) = 26.14$, $P = 0.0000042$). However, the effect on interaction was not significant. In control mice, there was significant main effect on time ($F(2,4) = 5.90$, $P = 0.064$), but no significant main effect on light ($F(1,2) = 2.7$, $P = 0.24$) and on interaction ($F(2,4) = 2.70$, $P = 0.181$).

These results indicate the opposite effect of DRN 5-HT photoinhibition on response time to choose and initiate a trial. The DRN 5-HT photoinhibition increased the time to initiate a trial, while inhibition promoted quicker decisions.

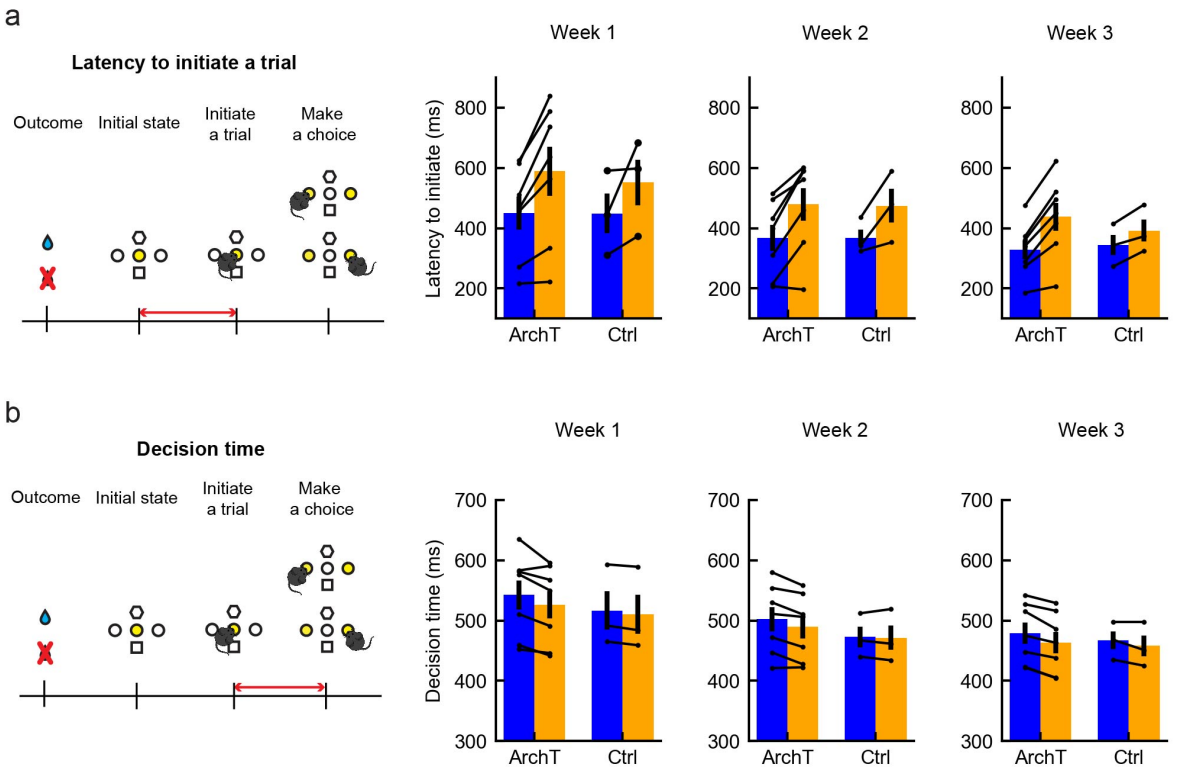

Figure 4.10: DRN 5-HT photoinhibition affected response times. a: The left drawing shows the definition of latency to initiate a trial. The right graphs indicate latency to initiate a trial in ArchT and control mice ($n = 7$ ArchT and 3 control mice) in each testing week. Heights of blue and orange bars indicate the mean across mice. b: The left drawing shows the definition of the time to make a first-step choice. Right-hand graphs indicate the time to make a first-step choice in ArchT and control mice ($n = 7$ ArchT and 3 control mice) in each testing week. Heights of blue and orange bars indicate the mean across mice in blue and yellow light stimulation, respectively. Error bars represent the SEM in all graphs.

4.4 Discussion

In this study, I examined the role of DRN 5-HT neurons in model-based decision making using a mouse behavioral task. I first showed that mice were most likely to use both model-free and model-based decision making. Based on the hypothesis that DRN 5-HT neurons control the weight of model-based decision making, I simulated the predicted effect of DRN 5-HT photoinhibition on choice behaviors. I then examined whether inhibition of DRN 5-HT neural activities affects behaviors related to first-step choice. The photoinhibition effect was partly consistent with the predicted effect in that stay probability after rare rewarded trials, which is indicative of model-based decision making, was increased. To quantitatively describe which computational process was affected by DRN 5-HT photoininhibition, I fitted the model-free/model-based RL model to behavioral data. The RL model analysis revealed that DRN 5-HT photoinhibition specifically affected the weight of model-based decision making. Lastly. I also examined how DRN 5-HT photoinhibition changed response time. Particularly, decreased time to make a decision possibly reflects the reduction of model-based decision making, as explained later in this section.

The two-step decision making task was initially developed for human subjects and has been used to examine how the brain arbitrates between model-free and model-based decision making [42]. Recently the task was also adapted for rodents, mainly in rats [68, 74, 119], but studies in mice are still limited [3]. In the present study, model evaluation showed that the model-free/model-based hybrid model could predict mouse choices the most accurately. However, improvement in model accuracy by the hybrid model from the model-based decision making model was relatively small, suggesting that mice largely rely on model-based decision making. This result is possibly due to the task structure. The two-step task in the present study included not only free-choice trials, but also forced-choice trials. These trials forced mice to sample information of action-state transitions. This might cause them to use the internal state transition.

Another point is that in 2 out of 6 mice, the inference model had the largest likelihood, as a recent previous study using the same configuration of two-step task suggested [17]. The former study reported that the inference model better fitted mouse behaviors than the hybrid model without forgetting. However, I also found that the hybrid RL model fit better than the inference model by adding the forgetting term. Although those two models describe computationally different processes, they are similar in that the agents update values of the first-step action that was not selected and the second-step state unattained. In the hybrid RL model, the value-forgetting process leads agents to update those values. On the other hand, the inference model updates them by using an anti-correlated structure of reward probabilities. As mentioned in the previous study, it is hard to dissociate these two decision making processes [2]. Therefore, it is important to find possible behavioral markers to dissociate them. Because in both strategies, subjects use the knowledge of state transition, it is difficult to dissociate them with stay probability analysis. The next important step to find a way to dissociate them is simulating behaviors of inference and model-based agents to find differences in their choice patterns.

The analysis with the hybrid RL model found that DRN 5-HT photoinhibition decreased the weight of model-based decision making. I also found that the balance

between model-free and model-based decision making was shifted toward model-free decision making by DRN 5-HT photoinhibition. This result was also consistent with previous studies in humans and mice [137, 187, 201]. In the human studies, they systemically reduced 5-HT level and examined the effect on choice behaviors in a two-step decision making task. By using rodents, the present study specifically examined DRN 5-HT neurons to identify more detailed neural substrates. With similar motivation, a recent study using a mouse reward devaluation task showed that DRN 5-HT inhibition increased habitual responses even after reward devaluation. However, the reward devaluation task has difficulty understanding detailed computational processes affected by DRN 5-HT inhibition. Increased habitual responses can be explained by increased use of model-free decision-making or by disrupted model-based decision making. In the present study, by using an RL model, I found that DRN 5-HT neurons specifically disrupted model-based decision making, but did not affect model-free decision making. The present study revealed a more detailed computational role of DRN 5-HT neurons.

I also found that DRN 5-HT photoinhibition affected response times. DRN photoinhibition increased the time to initiate a trial, while inhibition promoted a faster response in making the first-step decision. A possible reason for the change in response times is that photoinhibition affected the computational processes to guide actions. A previous electrophysiological study showed that background tonic modulation of DRN 5-HT neurons signals the state value of appetitive and aversive blocks of a Pavlovian conditioning task [29]. State values modulate the response vigor of actions [190]. In this study, photoinhibition of DRN 5-HT neurons may decrease the initial state value, resulting in longer latency to start a trial by photoinhibition.

On the other hand, inhibition decreased the time to make a first-step choice. Faster decisions might reflect decreased weight on model-based decision making. Model-based decision making requires calculating action values based on internal models, which is computationally expensive and is called a deliberative process [151]. On the other hand, model-free decision making does not entail so great a computational cost. A previous study using spatial navigation task showed that rats using place strategy displayed more vicarious trial-and-error, suggesting model-based strategy is underlying deliberative processes [66]. Another rat study showed that the deliberative process of model-based decision making could be observed as vicarious trial and error, such as pausing or head-turning at the point where decisions are needed [74]. The effect of photoinhibition on decision time may indicate that mice decided the first-step choice with less deliberation under photoinhibition. One alternative explanation is that decreased decision time reflects increased locomotor activity resulting from decreased 5-HT neural activity. Consistent with this explanation, previous studies showed that DRN 5-HT neural activity negatively controlled locomotion activity in an open-field test [32, 137, 164]. However, one study also showed that locomotion while mice were performing reward-driven motivated behaviors was not affected by optogenetic activation of DRN 5-HT neurons, suggesting that the role of DRN 5-HT neurons in locomotion is context-dependent [32]. In a two-step decision-making task, mice made decisions under motivation to obtain a reward, so the explanation that reduced decision time reflects increased locomotion speed is unlikely. A future experiment inhibiting DRN 5-HT neural activity before free-choice trials and forced-choice trials and comparing the effect on response time might allow us to clarify whether the change in locomotor

activity causes the decrease in decision time. Furthermore, based on control mouse data, it remains possible that yellow-light stimulation itself biases response time. It is necessary to examine this point by collecting more data from control mice. Contrary to the present study, another study using a probabilistic choice task did not show an optogenetic effect on decision time [63]. A possible reason for this discrepancy is the difference in decision-making systems to solve the tasks. A computational study modeling choice behaviors of Fonseca et al. [63] showed that mouse choice behaviors can be explained by a strategy which uses either win-stay lose-shift and model-free decision making, depending on how much time is spent to initiate the next trial [85]. This computational work suggests that the underlying decision-system Fonseca et al. [63] is different from that in the present study, in which mice mainly relied on model-based decision making. The role of DRN 5-HT neurons may be different in model-free and model-based decision making.

DRN 5-HT neurons project to diverse brain regions, and behavioral functions are complex. Which DRN 5-HT projections are responsible for value learning and arbitration of two valuation systems? One possibility is regulation of DA release in the striatum. Previous studies have examined the role of dopaminergic neurons from model-based value update to arbitration between model-free and model-based decision making (reviewed in [1]). DRN 5-HT neurons send projections to the VTA and control the release of DA in the ventral striatum via co-released glutamate [191]. During a task, DRN 5-HT-VTA projections may regulate DA release, affecting control of model-based decision making. In order to examine whether interaction with DA neurons is a possible mechanism to control model-based decision making, it will be the first step to measure DA activity induction with our stimulation protocol using neurochemical measurements such as *in vivo* microdialysis or optical imaging of DA biosensor signals. DRN projections to frontal regions such as the OFC and mPFC are also possible neural substrates. Previous studies in waiting tasks showed that model-based value estimates are sent to the OFC and mPFC and lead to increased waiting by DRN 5-HT activation [125]. In addition, lesioning the OFC decreased the weight of model-based decision making in a rat study using a two-step task [119]. Other studies found markers of model-based decision making in the OFC and other prefrontal regions [3, 54, 92, 185]. It is necessary to manipulate or monitor DRN 5-HT neural activities to specific brain regions to understand neural substrates further.

In conclusion, the present study demonstrated that DRN 5-HT neurons control reliance on model-based systems to select actions. These results showed that DRN 5-HT neurons specifically control the weight of model-based decision making, revealing a detailed computational role of DRN 5-HT neurons in model-based computation.

Chapter 5

Conclusion and Future directions

5.1 Summary of findings

The aim of this thesis was to understand the role of DRN 5-HT neurons in reward-based behaviors. Specifically, I examined two hypotheses on the role of 5-HT neurons based on the RL framework in regard to behavior: 1. 5-HT controls the discount factor and 2. 5-HT controls model-based decision making.

In the discount factor hypothesis, previous behavioral studies have consistently shown that DRN 5-HT neural activities are causally related to waiting for delayed rewards. To further examine whether DRN 5-HT neural activity controls enhancement of actions for future rewards, I tested the effect of optogenetic activation and inhibition of DRN 5-HT neurons on sustained motor actions for future rewards. The present study revealed different regulation of DRN 5-HT neurons between persistent motor actions and stationary waiting for future rewards. DRN 5-HT activation/inhibition promotes/suppresses patience to wait bidirectionally, while the same manipulation did not affect persistence to sustain activity but an effect on slowing down action vigor by optogenetic activation.

For the hypothesis on model-based decision making, computational studies proposed the involvement of DRN 5-HT neurons in facilitating model-based decision making, but direct behavioral evidence has been limited. I trained mice to perform a two-step decision making task and tested how optogenetic silencing of DRN 5-HT neurons influenced the computational process of model-based decision making. By applying a reinforcement learning model analysis of choice behaviors, I demonstrated that DRN 5-HT neurons regulate reliance on model-based systems in making a choice. These results revealed specific computational roles of 5-HT neurons in regulating model-based computations.

In order to further understand the discount factor hypothesis, I first examined the role of DRN 5-HT neurons in sustained motor actions. However, unexpectedly, I found dissociation between sustained motor actions and stationary waiting. Such results and other recent behavioral studies promote reconsideration of the hypothetical role of DRN 5-HT neurons and proposed another hypothesis that DRN 5-HT neurons promote model-based decision making. I examined this hypothesis and found consistent behavioral results. My present study suggests possible roles of DRN 5-HT neurons in model-based computation which can also explain the role of DRN 5-HT neurons in

waiting behaviors. However, the question of whether the results of sustained motor actions can be integrated to understand more comprehensively the role of the 5-HT system remains open. A recent study showed that sustained motor action is regulated by MRN 5-HT neurons, rather than those in the DRN [207]. Understanding how MRN 5-HT neurons interact with and modulate DRN 5-HT neurons during reward-based behaviors is an important next step to capture a comprehensive view of the role of 5-HT neurons in reward-based adaptive behaviors.

5.2 Limitations of the present study

In the present study, I took advantage of the optogenetic technique to selectively stimulate DRN 5-HT neural activity in a temporally precise manner. However, chronic long-term stimulation may cause damage to cells or reduced efficacy of optogenetic stimulation, especially for photoinhibition experiments [199]. It was reported that long-term stimulation of ArchT reduced the hyperpolarizing photocurrent [113]. In the present study, the duration of photostimulation was dependent on task performance of mice, which varied trial by trial from less than 1 minute to about 10 minutes. Therefore, it is possible that suppression efficacy was not constant during the experiments. Future behavioral experiments will need to monitor DRN 5-HT neural activities during behavioral tasks and will need to design a photostimulation protocol based on observed neural activities in order to effectively stimulate them. Alternative way to overcome the issues associated with ArchT is to use recently developed optogenetic tools. For example, light-gated chloride channels such as GtACR1 or GtACR2 could effectively inhibit neural activities with less light intensity and off-target effects, overcoming the issues associated with ion-pump based tools including ArchT [106]. Several other recent studies developed efficient optogenetic silencing tools for axon terminal inhibition such as lamprey paraopsin [31] and mosquito-based rhodopsin [107]. Using such tools would overcome decreased suppression efficacy and also enable to further investigate detailed neural substrates of behaviors.

RL model analysis clarifies the effect of DRN 5-HT photoinhibition on the computational process of decision making, but further validation of possible models may lead to a more precise understanding of the role of DRN 5-HT neurons in reward-driven decision making. In the present study, I used a model-based RL algorithm which has been commonly used in the literature. However, recent studies have proposed different frameworks that may also capture planning and flexible behaviors in two-step tasks, such as successor representation [156], meta-RL [192], and linear RL [140]. It is necessary to evaluate whether these models can better capture mouse behaviors in the two-step task.

5.3 Future directions

Sustained motor actions for future rewards: In the present study, different effects were observed in reward-based action sustainment between stationary waiting and motor actions. The most important question is, "How are sustained motor actions

regulated by the 5-HT system?" A previous study showed the importance of MRN 5-HT projections to the ventral hippocampus in regulation of sustainment of motivated behaviors [207]. Even within the DRN, there are multiple populations with distinct projection targets, such as DRN 5-HT-OFC or -mPFC projections, with different behavioral roles [152, 194]. By expressing ChR2 or ArchT in neurons with projection targets, it will be possible to test whether each 5-HT projection acts differently on sustained motor actions for future rewards. Another important future study is how sustained actions are regulated by DRN 5-HT neurons under the prediction of punishment or aversive stimuli. DRN 5-HT neurons encode not only reward, but also punishment signals, as previous studies have suggested [29, 110]. Examining how behaviors are regulated differently by reward and punishment will reveal a more complete picture of DRN 5-HT neurons in action sustainment for future outcomes.

Model-based decision making: An interesting future study is to address how task structure affects the effect of DRN 5-HT manipulation. A previous study in the stationary waiting task suggested that reward timing uncertainty modulates the effect of DRN 5-HT activation [124]. In the case of the two-step task, we can manipulate transition probability, reward probability, or the fraction of forced-choice task. How these changes affect choice behaviors and the effect of DRN 5-HT manipulation will allow a more precise interpretation of how animals tune their behaviors based on task structure and how 5-HT neurons are involved in the process. Also, whether DRN 5-HT neurons are involved in learning an internal model of the action-outcome relationship is an open research question.

Combination of neural manipulation with recording of neural activities: As a general future direction of my study, combining neural manipulation with recording, such as optical imaging or vivo electrophysiological recording, is desired. Thanks to the viral technique, it is possible to record from neurons in a cell-type-specific or pathway-specific manner. By injecting a viral vector containing calcium indicators such as GCaMP under specific promoters for 5-HT neurons, we can monitor neural activities of DRN 5-HT neurons during behavioral tasks. For the lever-pressing task, if there is modulation of neural activities at specific time points, e.g. before the onset of lever-press or just before abandoning a trial, I could design a photostimulation protocol to examine the role of such neural activities in sustained motor actions. Also, by labeling specific DRN 5-HT populations to a brain region, it is also possible to demonstrate the behavioral role of different projections. For the two-step task, by combining RL analysis of behaviors and functional recording experiments, neural correlates of latent decision variables can be examined as in previous studies [3, 87, 88]. To understand how DRN 5-HT neural activities functionally regulate model-based decision making, optical imaging or electrophysiological recording of 5-HT neurons during the two-step task will be necessary.

Toward understanding psychiatric disorders: Finally, the present study will also lead to further understanding of neural substrates of behavioral abnormalities in neuropsychiatric disorders. Sustained motor actions or effort expenditure for future rewards is an important behavioral characteristic of motivated behaviors. Disruption of such a cognitive process might be related to apathetic symptoms observed in multiple psychiatric disorders [9]. Also, disruption of model-based RL systems has been exam-

ined in patients with psychiatric disorders such as affective disorders [77], schizophrenia [38], and obsessive-compulsive disorders [186], as well as in healthy subjects under stress [33, 138]. Understanding neurochemical regulation of model-based decision making will lead us to better understand how disruption of model-based behavioral control occurs in psychiatric disorders.

Bibliography

[1] T. Akam and M. Walton. What is dopamine doing in model-based reinforcement learning? *Current Opinion in Behavioral Sciences*, 38:74–82, 2021.

[2] T. Akam, R. Costa, and P. Dayan. Simple plans or sophisticated habits? state, transition and learning interactions in the two-step task. *PLoS Comput Biol*, 11 (12):e1004648, 2015.

[3] T. Akam, I. Rodrigues-Vaz, I. Marcelo, X. Zhang, M. Pereira, R. F. Oliveira, P. Dayan, and R. M. Costa. The anterior cingulate cortex predicts future states to mediate model-based action selection. *Neuron*, 109(1):149–163 e7, 2021.

[4] K. D. Alex and E. A. Pehek. Pharmacologic mechanisms of serotonergic regulation of dopamine neurotransmission. *Pharmacol Ther*, 113(2):296–320, 2007.

[5] K. D. Alex, G. J. Yavanian, H. G. McFarlane, C. P. Pluto, and E. A. Pehek. Modulation of dopamine release by striatal 5-ht2c receptors. *Synapse*, 55(4): 242–51, 2005.

[6] J. Amat, M. V. Baratta, E. Paul, S. T. Bland, L. R. Watkins, and S. F. Maier. Medial prefrontal cortex determines how stressor controllability affects behavior and dorsal raphe nucleus. *Nat Neurosci*, 8(3):365–71, 2005.

[7] A. F. Arnsten and P. S. Goldman-Rakic. Selective prefrontal cortical projections to the region of the locus coeruleus and raphe nuclei in the rhesus monkey. *Brain Res*, 306(1-2):9–18, 1984.

[8] E. C. Azmitia. Evolution of serotonin: sunlight to suicide. In C. P. Muller and B. L. Jacobs, editors, *Handbook of the Behavioral Neurobiology of Serotonin*, chapter 1.1, pages 3–22. Academic Press, London, 2010.

[9] M. R. Bailey, E. H. Simpson, and P. D. Balsam. Neural substrates underlying effort, time, and risk-based decision making in motivated behavior. *Neurobiol Learn Mem*, 133:233–256, 2016.

[10] M. R. Bailey, C. Williamson, C. Mezias, V. Winiger, R. Silver, P. D. Balsam, and E. H. Simpson. The effects of pharmacological modulation of the serotonin 2c receptor on goal-directed behavior in mice. *Psychopharmacology (Berl)*, 233 (4):615–24, 2016.

[11] M. R. Bailey, O. Goldman, E. P. Bello, M. O. Chohan, N. Jeong, V. Winiger, E. Chun, E. Schipani, A. Kalmbach, J. F. Cheer, P. D. Balsam, and E. H. Simpson. An interaction between serotonin receptor signaling and dopamine enhances goal-directed vigor and persistence in mice. *J Neurosci*, 38(9):2149–2162, 2018.

[12] A. Bari, D. E. Theobald, D. Caprioli, A. C. Mar, A. Aidoo-Micah, J. W. Dalley, and T. W. Robbins. Serotonin modulates sensitivity to reward and negative feedback in a probabilistic reversal learning task in rats. *Neuropsychopharmacology*, 35(6):1290–301, 2010.

[13] R. L. Barlow, J. Alsio, B. Jupp, R. Rabinovich, S. Shrestha, A. C. Roberts, T. W. Robbins, and J. W. Dalley. Markers of serotonergic function in the orbitofrontal cortex and dorsal raphe nucleus predict individual variation in spatial-discrimination serial reversal learning. *Neuropsychopharmacology*, 40(7):1619–30, 2015.

[14] N. M. Barnes and T. Sharp. A review of central 5-ht receptors and their function. *Neuropharmacology*, 38(8):1083–152, 1999.

[15] G. Bezzina, S. Body, T. H. Cheung, C. L. Hampson, C. M. Bradshaw, J. C. Glennon, and E. Szabadi. Evidence for a role of 5-ht2c receptors in the motor aspects of performance, but not the efficacy of food reinforcers, in a progressive ratio schedule. *Psychopharmacology (Berl)*, 232(4):699–711, 2015.

[16] S. J. Bishop and C. Gagne. Anxiety, depression, and decision making: A computational perspective. *Annu Rev Neurosci*, 41:371–388, 2018.

[17] M. Blanco-Pozo, T. Akam, and M. Walton. Dopamine reports reward prediction errors, but does not update policy, during inference-guided choice. *bioRxiv*, 2021.

[18] M. Bortolato, K. Chen, and J. C. Shih. Monoamine oxidase inactivation: From pathophysiology to therapeutics. *Advanced Drug Delivery Reviews*, 60(13):1527–1533, 2008.

[19] Y. L. Boureau and P. Dayan. Opponency revisited: competition and cooperation between dopamine and serotonin. *Neuropsychopharmacology*, 36(1):74–97, 2011.

[20] C. M. Bradshaw and P. R. Killeen. A theory of behaviour on progressive ratio schedules, with applications in behavioural pharmacology. *Psychopharmacology (Berl)*, 222(4):549–64, 2012.

[21] E. S. Bromberg-Martin, O. Hikosaka, and K. Nakamura. Coding of task reward value in the dorsal raphe nucleus. *J Neurosci*, 30(18):6262–72, 2010.

[22] C. J. Browne, A. R. Abela, D. Chu, Z. Li, X. Ji, E. K. Lambe, and P. J. Fletcher. Dorsal raphe serotonin neurons inhibit operant responding for reward via inputs to the ventral tegmental area but not the nucleus accumbens: evidence from studies combining optogenetic stimulation and serotonin reuptake inhibition. *Neuropsychopharmacology*, 44(4):793–804, 2019.

[23] M. J. Bubar and K. A. Cunningham. Distribution of serotonin 5-ht2c receptors in the ventral tegmental area. *Neuroscience*, 146(1):286–97, 2007.

[24] M. J. Bubar, S. J. Stutz, and K. A. Cunningham. 5-ht(2c) receptors localize to dopamine and gaba neurons in the rat mesoaccumbens pathway. *PLoS One*, 6 (6):e20508, 2011.

[25] T. Canli and K. P. Lesch. Long story short: the serotonin transporter in emotion regulation and social cognition. *Nat Neurosci*, 10(9):1103–9, 2007.

[26] J. P. Chen, H. M. van Praag, and E. L. Gardner. Activation of 5-ht3 receptor by 1-phenylbiguanide increases dopamine release in the rat nucleus accumbens. *Brain Res*, 543(2):354–7, 1991.

[27] H. F. Clarke, J. W. Dalley, H. S. Crofts, T. W. Robbins, and A. C. Roberts. Cognitive inflexibility after prefrontal serotonin depletion. *Science*, 304(5672): 878–80, 2004.

[28] H. F. Clarke, S. C. Walker, J. W. Dalley, T. W. Robbins, and A. C. Roberts. Cognitive inflexibility after prefrontal serotonin depletion is behaviorally and neurochemically specific. *Cereb Cortex*, 17(1):18–27, 2007.

[29] J. Y. Cohen, M. W. Amoroso, and N. Uchida. Serotonergic neurons signal reward and punishment on multiple timescales. *Elife*, 4, 2015.

[30] R. Cools, K. Nakamura, and N. D. Daw. Serotonin and dopamine: unifying affective, activational, and decision functions. *Neuropsychopharmacology*, 36(1): 98–113, 2011.

[31] B. A. Copits, R. Gowrishankar, P. R. O'Neill, J.-N. Li, K. S. Girven, J. J. Yoo, X. Meshik, K. E. Parker, S. M. Spangler, A. J. Elerding, B. J. Brown, S. E. Shirley, K. K. L. Ma, A. M. Vasquez, M. C. Stander, V. Kalyanaraman, S. K. Vogt, V. K. Samineni, T. Patriarchi, L. Tian, N. Gautam, R. K. Sunahara, R. W. I. V. Gereau, and M. R. Bruchas. A photoswitchable gpcr-based opsin for presynaptic inhibition. *Neuron*, 109(11):1791–1809.e11, 2021.

[32] P. A. Correia, E. Lottem, D. Banerjee, A. S. Machado, M. R. Carey, and Z. F. Mainen. Transient inhibition and long-term facilitation of locomotion by phasic optogenetic activation of serotonin neurons. *Elife*, 6, 2017.

[33] A. Cremer, F. Kalbe, J. Glascher, and L. Schwabe. Stress reduces both model-based and model-free neural computations during flexible learning. *Neuroimage*, 229:117747, 2021.

[34] M. J. Crockett and R. Cools. Serotonin and aversive processing in affective and social decision-making. *Current Opinion in Behavioral Sciences*, 5:64–70, 2015.

[35] M. J. Crockett, L. Clark, and T. W. Robbins. Reconciling the role of serotonin in behavioral inhibition and aversion: acute tryptophan depletion abolishes punishment-induced inhibition in humans. *J Neurosci*, 29(38):11993–9, 2009.

[36] M. J. Crockett, L. Clark, A. M. Apergis-Schoute, S. Morein-Zamir, and T. W. Robbins. Serotonin modulates the effects of pavlovian aversive predictions on response vigor. *Neuropsychopharmacology*, 37(10):2244–52, 2012.

[37] D. A. Cruz, S. M. Eggan, E. C. Azmitia, and D. A. Lewis. Serotonin1a receptors at the axon initial segment of prefrontal pyramidal neurons in schizophrenia. *Am J Psychiatry*, 161(4):739–42, 2004.

[38] A. J. Culbreth, A. Westbrook, N. D. Daw, M. Botvinick, and D. M. Barch. Reduced model-based decision-making in schizophrenia. *J Abnorm Psychol*, 125 (6):777–787, 2016.

[39] A. Czyrak, K. Czepiel, M. Mackowiak, A. Chocyk, and K. Wedzony. Serotonin 5-ht1a receptors might control the output of cortical glutamatergic neurons in rat cingulate cortex. *Brain Res*, 989(1):42–51, 2003.

[40] N. D. Daw, S. Kakade, and P. Dayan. Opponent interactions between serotonin and dopamine. *Neural Netw*, 15(4-6):603–16, 2002.

[41] N. D. Daw, Y. Niv, and P. Dayan. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci*, 8 (12):1704–11, 2005.

[42] N. D. Daw, S. J. Gershman, B. Seymour, P. Dayan, and R. J. Dolan. Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69 (6):1204–15, 2011.

[43] P. Dayan and Q. J. Huys. Serotonin, inhibition, and negative mood. *PLoS Comput Biol*, 4(2):e4, 2008.

[44] P. Dayan and Q. J. Huys. Serotonin in affective control. *Annu Rev Neurosci*, 32: 95–126, 2009.

[45] P. De Deurwaerdere, L. Stinus, and U. Spampinato. Opposite change of in vivo dopamine release in the rat nucleus accumbens and striatum that follows electrical stimulation of dorsal raphe nucleus: role of 5-ht3 receptors. *J Neurosci*, 18 (16):6528–38, 1998.

[46] J. F. Deakin and F. G. Graeff. 5-ht and mechanisms of defence. *J Psychopharmacol*, 5(4):305–15, 1991.

[47] F. Denk, M. E. Walton, K. A. Jennings, T. Sharp, M. F. Rushworth, and D. M. Bannerman. Differential involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort. *Psychopharmacology (Berl)*, 179(3):587–96, 2005.

[48] R. J. Dolan and P. Dayan. Goals and habits in the brain. *Neuron*, 80(2):312–25, 2013.

[49] B. B. Doll, D. A. Simon, and N. D. Daw. The ubiquity of model-based reinforcement learning. *Curr Opin Neurobiol*, 22(6):1075–81, 2012.

[50] K. Doya. Metalearning and neuromodulation. *Neural Netw*, 15(4-6):495–506, 2002.

[51] K. Doya. Modulators of decision making. *Nat Neurosci*, 11(4):410–6, 2008.

[52] K. Eberle-Wang, Z. Mikeladze, K. Uryu, and M. F. Chesselet. Pattern of expression of the serotonin2c receptor messenger rna in the basal ganglia of adult rats. *J Comp Neurol*, 384(2):233–47, 1997.

[53] I. Ferezou, B. Cauli, E. L. Hill, J. Rossier, E. Hamel, and B. Lambolez. 5-ht3 receptors mediate serotonergic fast synaptic excitation of neocortical vasoactive intestinal peptide/cholecystokinin interneurons. *J Neurosci*, 22(17):7389–97, 2002.

[54] A. S. Fermin, T. Yoshida, J. Yoshimoto, M. Ito, S. C. Tanaka, and K. Doya. Model-based action planning involves cortico-cerebellar and basal ganglia networks. *Sci Rep*, 6:31378, 2016.

[55] M. Filip and M. Bader. Overview on 5-ht receptors and their role in physiology and pathology of the central nervous system. *Pharmacol Rep*, 61(5):761–77, 2009.

[56] M. Filip, U. Spampinato, A. C. McCreary, and E. Przegalinski. Pharmacological and genetic interventions in serotonin (5-ht)(2c) receptors to alter drug abuse and dependence processes. *Brain Res*, 1476:132–53, 2012.

[57] P. J. Fletcher. Effects of d-fenfluramine and metergoline on responding for conditioned reward and the response potentiating effect of nucleus accumbens d-amphetamine. *Psychopharmacology (Berl)*, 118(2):155–63, 1995.

[58] P. J. Fletcher. Injection of 5-ht into the nucleus accumbens reduces the effects of d-amphetamine on responding for conditioned reward. *Psychopharmacology (Berl)*, 126(1):62–9, 1996.

[59] P. J. Fletcher and K. M. Korth. Activation of 5-ht1b receptors in the nucleus accumbens reduces amphetamine-induced enhancement of responding for conditioned reward. *Psychopharmacology (Berl)*, 142(2):165–74, 1999.

[60] P. J. Fletcher, A. Azampanah, and K. M. Korth. Activation of 5-ht(1b) receptors in the nucleus accumbens reduces self-administration of amphetamine on a progressive ratio schedule. *Pharmacol Biochem Behav*, 71(4):717–25, 2002.

[61] P. J. Fletcher, A. F. Chintoh, J. Sinyard, and G. A. Higgins. Injection of the 5-ht2c receptor agonist ro60-0175 into the ventral tegmental area reduces cocaine-induced locomotor activity and cocaine self-administration. *Neuropsychopharmacology*, 29(2):308–18, 2004.

[62] P. J. Fletcher, Z. Rizos, K. Noble, A. D. Soko, L. B. Silenieks, A. D. Le, and G. A. Higgins. Effects of the 5-ht2c receptor agonist ro60-0175 and the 5-ht2a receptor antagonist m100907 on nicotine self-administration and reinstatement. *Neuropharmacology*, 62(7):2288–98, 2012.

[63] M. S. Fonseca, M. Murakami, and Z. F. Mainen. Activation of dorsal raphe serotonergic neurons promotes waiting but is not reinforcing. *Curr Biol*, 25(3): 306–15, 2015.

[64] K. B. J. Franklin and G. Paxinos. *The Mouse Brain in Stereotaxic Coordinates compact 3rd edition*. Academic Press, 2008.

[65] W. Fu, E. Le Maitre, V. Fabre, J. F. Bernard, Z. Q. David Xu, and T. Hokfelt. Chemical neuroanatomy of the dorsal raphe nucleus and adjacent structures of the mouse brain. *J Comp Neurol*, 518(17):3464–94, 2010.

[66] R. S. Gardner, M. R. Uttaro, S. E. Fleming, D. F. Suarez, G. A. Ascoli, and T. C. Dumas. A secondary working memory challenge preserves primary place strategies despite overtraining. *Learn Mem*, 20(11):648–56, 2013.

[67] H. J. Groenewegen and H. B. Uylings. The prefrontal cortex and the integration of sensory, limbic and autonomic information. *Prog Brain Res*, 126:3–28, 2000.

[68] S. M. Groman, B. Massi, S. R. Mathias, D. W. Curry, D. Lee, and J. R. Taylor. Neurochemical and behavioral dissections of decision-making in a rodent multistage task. *J Neurosci*, 39(2):295–306, 2019.

[69] C. D. Grossman, B. A. Bari, and J. Y. Cohen. Serotonin neurons modulate learning rate through uncertainty. *Current Biology*, 2021.

[70] M. Hajos, S. E. Gartside, V. Varga, and T. Sharp. In vivo inhibition of neuronal activity in the rat ventromedial prefrontal cortex by midbrain-raphe nuclei: role of 5-ht1a receptors. *Neuropharmacology*, 45(1):72–81, 2003.

[71] A. A. Harrison, B. J. Everitt, and T. W. Robbins. Central 5-ht depletion enhances impulsive responding without affecting the accuracy of attentional performance: interactions with dopaminergic mechanisms. *Psychopharmacology (Berl)*, 133(4): 329–42, 1997.

[72] S. Hashimoto, T. Inoue, and T. Koyama. Serotonin reuptake inhibitors reduce conditioned fear stress-induced freezing behavior in rats. *Psychopharmacology (Berl)*, 123(2):182–6, 1996.

[73] S. Hashimoto, T. Inoue, and T. Koyama. Effects of conditioned fear stress on serotonin neurotransmission and freezing behavior in rats. *Eur J Pharmacol*, 378 (1):23–30, 1999.

[74] B. M. Hasz and A. D. Redish. Deliberation and procedural automation on a two-step task for rats. *Front Integr Neurosci*, 12:30, 2018.

[75] K. Hayashi, K. Nakao, and K. Nakamura. Appetitive and aversive information coding in the primate dorsal raphe nucleus. *J Neurosci*, 35(15):6195–208, 2015.

[76] D. J. Hayes and A. J. Greenshaw. 5-ht receptors and reward-related behaviour: a review. *Neurosci Biobehav Rev*, 35(6):1419–49, 2011.

[77] S. Heo, Y. Sung, and S. W. Lee. Effects of subclinical depression on prefrontal-striatal model-based and model-free learning. *PLoS Comput Biol*, 17(5):e1009003, 2021.

[78] D. Herve, V. M. Pickel, T. H. Joh, and A. Beaudet. Serotonin axon terminals in the ventral tegmental area of the rat: fine structure and synaptic input to dopaminergic neurons. *Brain Res*, 435(1-2):71–83, 1987.

[79] G. A. Higgins, L. B. Silenieks, A. Rossmann, Z. Rizos, K. Noble, A. D. Soko, and P. J. Fletcher. The 5-ht2c receptor agonist lorcaserin reduces nicotine self-administration, discrimination, and reinstatement: relationship to feeding behavior and impulse control. *Neuropsychopharmacology*, 37(5):1177–91, 2012.

[80] G. A. Higgins, L. B. Silenieks, W. Lau, I. A. de Lannoy, D. K. Lee, J. Izhakova, K. Coen, A. D. Le, and P. J. Fletcher. Evaluation of chemically diverse 5-ht(2)c receptor agonists on behaviours motivated by food and nicotine and on side effect profiles. *Psychopharmacology (Berl)*, 226(3):475–90, 2013.

[81] H. Hioki, H. Nakamura, Y. F. Ma, M. Konno, T. Hayakawa, K. C. Nakamura, F. Fujiyama, and T. Kaneko. Vesicular glutamate transporter 3-expressing non-serotonergic projection neurons constitute a subregion in the rat midbrain raphe nuclei. *J Comp Neurol*, 518(5):668–86, 2010.

[82] M. Y. Ho, S. Body, S. Kheramin, C. M. Bradshaw, and E. Szabadi. Effects of 8-oh-dpat and way-100635 on performance on a time-constrained progressive-ratio schedule. *Psychopharmacology (Berl)*, 167(2):137–44, 2003.

[83] J. R. Homberg. Serotonin and decision making processes. *Neurosci Biobehav Rev*, 36(1):218–36, 2012.

[84] D. Hoyer, J. P. Hannon, and G. R. Martin. Molecular, pharmacological and functional diversity of 5-ht receptors. *Pharmacol Biochem Behav*, 71(4):533–54, 2002.

[85] K. Iigaya, M. S. Fonseca, M. Murakami, Z. F. Mainen, and P. Dayan. An effect of serotonergic stimulation on learning rates for rewards apparent after long intertrial intervals. *Nat Commun*, 9(1):2477, 2018.

[86] K. Inaba, T. Mizuhiki, T. Setogawa, K. Toda, B. J. Richmond, and M. Shidara. Neurons in monkey dorsal raphe nucleus code beginning and progress of step-by-step schedule, reward expectation, and amount of reward outcome in the reward schedule task. *J Neurosci*, 33(8):3477–91, 2013.

[87] M. Ito and K. Doya. Validation of decision-making models and analysis of decision variables in the rat basal ganglia. *J Neurosci*, 29(31):9861–74, 2009.

[88] M. Ito and K. Doya. Distinct neural representation in the dorsolateral, dorsomedial, and ventral parts of the striatum during fixed- and free-choice tasks. *J Neurosci*, 35(8):3499–514, 2015.

[89] A. Izquierdo, K. Carlos, S. Ostrander, D. Rodriguez, A. McCall-Craddolph, G. Yagnik, and F. Zhou. Impaired reward learning and intact motivation after serotonin depletion in rats. *Behav Brain Res*, 233(2):494–9, 2012.

[90] B. L. Jacobs and E. C. Azmitia. Structure and function of the brain serotonin system. *Physiol Rev*, 72(1):165–229, 1992.

[91] R. L. Jakab and P. S. Goldman-Rakic. 5-hydroxytryptamine2a serotonin receptors in the primate cerebral cortex: possible site of action of hallucinogenic and antipsychotic drugs in pyramidal cell apical dendrites. *Proc Natl Acad Sci U S A*, 95(2):735–40, 1998.

[92] J. L. Jones, G. R. Esber, M. A. McDannald, A. J. Gruber, A. Hernandez, A. Mirenzi, and G. Schoenbaum. Orbitofrontal cortex supports behavior and learning using inferred but not cached values. *Science*, 338(6109):953–6, 2012.

[93] J. W. Kanen, A. M. Apergis-Schoute, R. Yellowlees, F. E. Arntz, F. E. van der Flier, A. Price, R. N. Cardinal, D. M. Christmas, L. Clark, B. J. Sahakian, M. J. Crockett, and T. W. Robbins. Serotonin depletion impairs both pavlovian and instrumental reversal learning in healthy humans. *Mol Psychiatry*, 2021.

[94] P. R. Killeen. Mathematical principles of reinforcement. *Behav and Brain Sci*, 17:105–172, 1994.

[95] P. R. Killeen, D. Posadas-Sanchez, E. B. Johansen, and E. A. Thrailkill. Progressive ratio schedules of reinforcement. *J Exp Psychol Anim Behav Process*, 35 (1):35–50, 2009.

[96] A. R. Kosheleff, M. Grimes, S. J. O'Dell, J. F. Marshall, and A. Izquierdo. Work aversion and associated changes in dopamine and serotonin transporter after methamphetamine exposure in rats. *Psychopharmacology (Berl)*, 219(2):411–20, 2012.

[97] A. M. Kulichenko and V. B. Pavlenko. Self-initiated behavioral act-related neuronal acitivity in the region of the raphe nuclei of the cat. *Neurophysiology*, 36 (1):50–57, 2004.

[98] D. Lee. Decision making: from neuroscience to psychiatry. *Neuron*, 78(2):233–48, 2013.

[99] Y. Li, W. Zhong, D. Wang, Q. Feng, Z. Liu, J. Zhou, C. Jia, F. Hu, J. Zeng, Q. Guo, L. Fu, and M. Luo. Serotonin neurons in the dorsal raphe nucleus encode reward signals. *Nat Commun*, 7:10503, 2016.

[100] R. Lin, J. Liang, R. Wang, T. Yan, Y. Zhou, Y. Liu, Q. Feng, F. Sun, Y. Li, A. Li, H. Gong, and M. Luo. The raphe dopamine system controls the expression of incentive memory. *Neuron*, 106(3):498–514 e8, 2020.

[101] R. Liu, T. Jolas, and G. Aghajanian. Serotonin 5-ht(2) receptors activate local gaba inhibitory inputs to serotonergic neurons of the dorsal raphe nucleus. *Brain Res*, 873(1):34–45, 2000.

[102] Z. Liu, J. Zhou, Y. Li, F. Hu, Y. Lu, M. Ma, Q. Feng, J. E. Zhang, D. Wang, J. Zeng, J. Bao, J. Y. Kim, Z. F. Chen, S. El Mestikawy, and M. Luo. Dorsal raphe neurons signal reward through 5-ht and glutamate. *Neuron*, 81(6):1360–74, 2014.

[103] Z. Liu, R. Lin, and M. Luo. Reward contributions to serotonergic functions. *Annu Rev Neurosci*, 43:141–162, 2020.

[104] E. Lottem, D. Banerjee, P. Vertechi, D. Sarra, M. O. Lohuis, and Z. F. Mainen. Activation of serotonin neurons promotes active persistence in a probabilistic foraging task. *Nat Commun*, 9(1):1000, 2018.

[105] M. Luo, Y. Li, and W. Zhong. Do dorsal raphe 5-ht neurons encode "beneficialness"? *Neurobiol Learn Mem*, 135:40–49, 2016.

[106] M. Mahn, L. Gibor, P. Patil, K. Cohen-Kashi Malina, S. Oring, Y. Printz, R. Levy, I. Lampl, and O. Yizhar. High-efficiency optogenetic silencing with soma-targeted anion-conducting channelrhodopsins. *Nature Communications*, 9 (1):4125, 2018.

[107] M. Mahn, I. Saraf-Sinik, P. Patil, M. Pulin, E. Bitton, N. Karalis, F. Bruentgens, S. Palgi, A. Gat, J. Dine, J. Wietek, I. Davidi, R. Levy, A. Litvin, F. Zhou, K. Sauter, P. Soba, D. Schmitz, A. Lüthi, B. R. Rost, J. S. Wiegert, and O. Yizhar. Efficient optogenetic silencing of neurotransmitter release with a mosquito rhodopsin. *Neuron*, 109(10):1621–1635.e8, 2021.

[108] G. J. Marek and G. K. Aghajanian. 5-ht2a receptor or alpha1-adrenoceptor activation induces excitatory postsynaptic currents in layer v pyramidal cells of the medial prefrontal cortex. *Eur J Pharmacol*, 367(2-3):197–206, 1999.

[109] R. Martin-Ruiz, M. V. Puig, P. Celada, D. A. Shapiro, B. L. Roth, G. Mengod, and F. Artigas. Control of serotonergic function in medial prefrontal cortex by serotonin-2a receptors through a glutamate-dependent mechanism. *J Neurosci*, 21(24):9856–66, 2001.

[110] S. Matias, E. Lottem, G. P. Dugue, and Z. F. Mainen. Activity patterns of serotonin neurons underlying cognitive flexibility. *Elife*, 6, 2017.

[111] S. Matthes and M. Bader. Peripheral serotonin synthesis as a new drug target. *Trends in Pharmacological Sciences*, 39(6):560–572, 2018.

[112] G. A. Matthews, E. H. Nieh, C. M. Vander Weele, S. A. Halbert, R. V. Pradhan, A. S. Yosafat, G. F. Glober, E. M. Izadmehr, R. E. Thomas, G. D. Lacy, C. P. Wildes, M. A. Ungless, and K. M. Tye. Dorsal raphe dopamine neurons represent the experience of social isolation. *Cell*, 164(4):617–31, 2016.

[113] J. Mattis, K. M. Tye, E. A. Ferenczi, C. Ramakrishnan, D. J. O'Shea, R. Prakash, L. A. Gunaydin, M. Hyun, L. E. Fenno, V. Gradinaru, O. Yizhar, and K. Deisseroth. Principles for applying optogenetic tools derived from direct comparative analysis of microbial opsins. *Nat Methods*, 9(2):159–72, 2011.

[114] D. McCarthy, P. Voss, and M. Davison. Leaving patches: Effects of travel requirements. *Journal of the experimental analysis of behavior*, 62:185–200, 10 1994.

[115] A. McGregor, S. Lacosta, and D. C. Roberts. L-tryptophan decreases the breaking point under a progressive ratio schedule of intravenous cocaine reinforcement in the rat. *Pharmacol Biochem Behav*, 44(3):651–5, 1993.

[116] J. Metcalfe and W. Mischel. A hot/cool-system analysis of delay of gratification: dynamics of willpower. *Psychol Rev*, 106(1):3–19, 1999.

[117] F. Meyniel, G. M. Goodwin, J. W. Deakin, C. Klinge, C. MacFadyen, H. Milligan, E. Mullings, M. Pessiglione, and R. Gaillard. A specific role for serotonin in overcoming effort cost. *Elife*, 5, 2016.

[118] J. Michely, E. Eldar, I. M. Martin, and R. J. Dolan. A mechanistic account of serotonin's impact on mood. *Nat Commun*, 11(1):2335, 2020.

[119] K. J. Miller, M. M. Botvinick, and C. D. Brody. Dorsal hippocampus contributes to model-based planning. *Nat Neurosci*, 20(9):1269–1276, 2017.

[120] W. Mischel and E. B. Ebbesen. Attention in delay of gratification. *Journal of Personality and Social Psychology*, 16(2):329, 1970.

[121] W. Mischel, A. R. Zeiss, and E. B. Ebbesen. Cognitive and attentional mechanisms in delay of gratification. *Journal of Personality and Social Psychology*, 21 (2):204, 1972.

[122] K. Miyazaki, K. W. Miyazaki, and K. Doya. Activation of dorsal raphe serotonin neurons underlies waiting for delayed rewards. *J Neurosci*, 31(2):469–79, 2011.

[123] K. Miyazaki, K. W. Miyazaki, and K. Doya. The role of serotonin in the regulation of patience and impulsivity. *Mol Neurobiol*, 45(2):213–24, 2012.

[124] K. Miyazaki, K. W. Miyazaki, A. Yamanaka, T. Tokuda, K. F. Tanaka, and K. Doya. Reward probability and timing uncertainty alter the effect of dorsal raphe serotonin neurons on patience. *Nat Commun*, 9(1):2048, 2018.

[125] K. Miyazaki, K. W. Miyazaki, G. Sivori, A. Yamanaka, K. F. Tanaka, and K. Doya. Serotonergic projections to the orbitofrontal and medial prefrontal cortices differentially modulate waiting for future rewards. *Sci Adv*, 6(48), 2020.

[126] K. W. Miyazaki, K. Miyazaki, and K. Doya. Activation of the central serotonergic system in response to delayed but not omitted rewards. *Eur J Neurosci*, 33(1): 153–60, 2011.

[127] K. W. Miyazaki, K. Miyazaki, and K. Doya. Activation of dorsal raphe serotonin neurons is necessary for waiting for delayed rewards. *J Neurosci*, 32(31):10451–7, 2012.

[128] K. W. Miyazaki, K. Miyazaki, K. F. Tanaka, A. Yamanaka, A. Takahashi, S. Tabuchi, and K. Doya. Optogenetic activation of dorsal raphe serotonin neurons enhances patience for future rewards. *Curr Biol*, 24(17):2033–40, 2014.

[129] M. Morales and F. E. Bloom. The 5-ht3 receptor is present in different subpopulations of gabaergic neurons in the rat telencephalon. *J Neurosci*, 17(9):3157–67, 1997.

[130] R. J. Moran, K. T. Kishida, T. Lohrenz, I. Saez, A. W. Laxton, M. R. Witcher, S. B. Tatter, T. L. Ellis, P. E. Phillips, P. Dayan, and P. R. Montague. The protective action encoding of serotonin transients in the human brain. *Neuropsychopharmacology*, 43(6):1425–1435, 2018.

[131] Y. Nagai, K. Takayama, N. Nishitani, C. Andoh, M. Koda, H. Shirakawa, T. Nakagawa, K. Nagayasu, A. Yamanaka, and S. Kaneko. The role of dorsal raphe serotonin neurons in the balance between reward and aversion. *Int J Mol Sci*, 21 (6), 2020.

[132] K. Nakamura, M. Matsumoto, and O. Hikosaka. Reward-dependent modulation of neuronal activity in the primate dorsal raphe nucleus. *J Neurosci*, 28(20): 5331–43, 2008.

[133] S. Navailles, D. Moison, K. A. Cunningham, and U. Spampinato. Differential regulation of the mesoaccumbens dopamine circuit by serotonin2c receptors in the ventral tegmental area and the nucleus accumbens: an in vivo microdialysis study with cocaine. *Neuropsychopharmacology*, 33(2):237–46, 2008.

[134] L. J. Nonkes, K. Tomson, A. Maertin, J. Dederen, J. H. Maes, and J. Homberg. Orbitofrontal cortex and amygdalar over-activity is associated with an inability to use the value of expected outcomes to guide behaviour in serotonin transporter knockout rats. *Neurobiol Learn Mem*, 94(1):65–72, 2010.

[135] S. K. Ogawa, J. Y. Cohen, D. Hwang, N. Uchida, and M. Watabe-Uchida. Organization of monosynaptic inputs to the serotonin and dopamine neuromodulatory systems. *Cell Rep*, 8(4):1105–18, 2014.

[136] Y. Ohmura, I. Tsutsui-Kimura, H. Sasamori, M. Nebuka, N. Nishitani, K. F. Tanaka, A. Yamanaka, and M. Yoshioka. Different roles of distinct serotonergic pathways in anxiety-like behavior, antidepressant-like, and anti-impulsive effects. *Neuropharmacology*, 167:107703, 2020.

[137] Y. Ohmura, K. Iwami, S. Chowdhury, H. Sasamori, C. Sugiura, Y. Bouchekioua, N. Nishitani, A. Yamanaka, and M. Yoshioka. Disruption of model-based decision making by silencing of serotonin neurons in the dorsal raphe nucleus. *Curr Biol*, 31(11):2446–2454 e5, 2021.

[138] A. R. Otto, C. M. Raio, A. Chiang, E. A. Phelps, and N. D. Daw. Working-memory capacity protects model-based learning from stress. *Proc Natl Acad Sci U S A*, 110(52):20941–6, 2013.

[139] C. Peyron, J. M. Petit, C. Rampon, M. Jouvet, and P. H. Luppi. Forebrain afferents to the rat dorsal raphe nucleus demonstrated by retrograde and anterograde tracing methods. *Neuroscience*, 82(2):443–68, 1998.

[140] P. Piray and N. D. Daw. Linear reinforcement learning in planning, grid fields, and cognitive control. *Nat Commun*, 12(1):4942, 2021.

[141] I. Pollak Dorocic, D. Furth, Y. Xuan, Y. Johansson, L. Pozzi, G. Silberberg, M. Carlen, and K. Meletis. A whole-brain atlas of inputs to serotonergic neurons of the dorsal and median raphe nuclei. *Neuron*, 83(3):663–78, 2014.

[142] S. Pommer, Y. Akamine, S. N. Schiffmann, A. de Kerchove d'Exaerde, and J. R. Wickens. The effect of serotonin receptor 5-ht1b on lateral inhibition between spiny projection neurons in the mouse striatum. *Journal of Neuroscience*, 41(37): 7831–7847, 2021.

[143] M. Pompeiano, J. M. Palacios, and G. Mengod. Distribution and cellular localization of mrna coding for 5-ht1a receptor in the rat brain: correlation with receptor binding. *J Neurosci*, 12(2):440–53, 1992.

[144] M. Pompeiano, J. M. Palacios, and G. Mengod. Distribution of the serotonin 5-ht2 receptor family mrnas: comparison between 5-ht2a and 5-ht2c receptors. *Brain Res Mol Brain Res*, 23(1-2):163–78, 1994.

[145] W. E. Pratt, M. A. Schall, and E. Choi. Selective serotonin receptor stimulation of the medial nucleus accumbens differentially affects appetitive motivation for food on a progressive ratio schedule of reinforcement. *Neurosci Lett*, 511(2):84–8, 2012.

[146] M. V. Puig and A. T. Gulledge. Serotonin and prefrontal cortex function: Neurons, networks, and circuits. *Molecular Neurobiology*, 44(3):449–464, 2011.

[147] M. V. Puig, F. Artigas, and P. Celada. Modulation of the activity of pyramidal neurons in rat prefrontal cortex by raphe stimulation in vivo: involvement of serotonin and gaba. *Cereb Cortex*, 15(1):1–14, 2005.

[148] M. V. Puig, A. Watakabe, M. Ushimaru, T. Yamamori, and Y. Kawaguchi. Serotonin modulates fast-spiking interneuron and synchronous activity in the rat prefrontal cortex through 5-ht1a and 5-ht2a receptors. *J Neurosci*, 30(6):2211–22, 2010.

[149] J. Qi, S. Zhang, H. L. Wang, H. Wang, J. de Jesus Aceves Buendia, A. F. Hoffman, C. R. Lupica, R. P. Seal, and M. Morales. A glutamatergic reward input from the dorsal raphe to ventral tegmental area dopamine neurons. *Nat Commun*, 5:5390, 2014.

[150] S. P. Ranade and Z. F. Mainen. Transient firing of dorsal raphe neurons encodes diverse and specific sensory, motor, and reward events. *J Neurophysiol*, 102(5): 3026–37, 2009.

[151] A. D. Redish. Vicarious trial and error. *Nature Reviews Neuroscience*, 17(3): 147–159, 2016.

[152] J. Ren, D. Friedmann, J. Xiong, C. D. Liu, B. R. Ferguson, T. Weerakkody, K. E. DeLoach, C. Ran, A. Pun, Y. Sun, B. Weissbourd, R. L. Neve, J. Huguenard, M. A. Horowitz, and L. Luo. Anatomically defined and functionally distinct dorsal raphe serotonin sub-systems. *Cell*, 175(2):472–487 e20, 2018.

[153] J. Ren, A. Isakova, D. Friedmann, J. Zeng, S. M. Grutzner, A. Pun, G. Q. Zhao, S. S. Kolluru, R. Wang, R. Lin, P. Li, A. Li, J. L. Raymond, Q. Luo, M. Luo, S. R. Quake, and L. Luo. Single-cell transcriptomes and whole-brain projections of serotonin neurons in the mouse dorsal and median raphe nuclei. *Elife*, 8, 2019.

[154] N. R. Richardson and D. C. Roberts. Progressive ratio schedules in drug self-administration studies in rats: a method to evaluate reinforcing efficacy. *J Neurosci Methods*, 66(1):1–11, 1996.

[155] B. A. Rocha, K. Scearce-Levie, J. J. Lucas, N. Hiroi, N. Castanon, J. C. Crabbe, E. J. Nestler, and R. Hen. Increased vulnerability to cocaine in mice lacking the serotonin-1b receptor. *Nature*, 393(6681):175–8, 1998.

[156] E. M. Russek, I. Momennejad, M. M. Botvinick, S. J. Gershman, and N. D. Daw. Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Comput Biol*, 13(9):e1005768, 2017.

[157] A. C. Sanders, A. J. Hussain, R. Hen, and X. Zhuang. Chronic blockade or constitutive deletion of the serotonin transporter reduces operant responding for food reward. *Neuropsychopharmacology*, 32(11):2321–9, 2007.

[158] N. Santana, A. Bortolozzi, J. Serrats, G. Mengod, and F. Artigas. Expression of serotonin1a and serotonin2a receptors in pyramidal and gabaergic neurons of the rat prefrontal cortex. *Cereb Cortex*, 14(10):1100–9, 2004.

[159] J. Scholl, N. Kolling, N. Nelissen, M. Browning, M. F. Rushworth, and C. J. Harmer. Beyond negative valence: 2-week administration of a serotonergic antidepressant enhances both reward and effort learning signals. *PLoS Biol*, 15(2): e2000756, 2017.

[160] J. H. Schwartz and J. A. Javitch. Neurotransmitters. In E. R. Kandel, J. H. Schwartz, T. Jessell, S. A. Siegelbaum, and A. J. Hudspeth, editors, *Principles of Neural Science*, book section 13, pages 289–306. McGraw-Hill, USA, 2013.

[161] N. Schweighofer, S. C. Tanaka, and K. Doya. Serotonin and the evaluation of future rewards: theory, experiments, and possible neural mechanisms. *Ann N Y Acad Sci*, 1104:289–300, 2007.

[162] N. Schweighofer, M. Bertin, K. Shishida, Y. Okamoto, S. C. Tanaka, S. Yamawaki, and K. Doya. Low-serotonin levels increase delayed reward discounting in humans. *J Neurosci*, 28(17):4528–32, 2008.

[163] A. Sengupta, M. Bocchio, D. M. Bannerman, T. Sharp, and M. Capogna. Control of amygdala circuits by 5-ht neurons via 5-ht and glutamate cotransmission. *J Neurosci*, 37(7):1785–1796, 2017.

[164] C. Seo, A. Guru, M. Jin, B. Ito, B. J. Sleezer, Y. Y. Ho, E. Wang, C. Boada, N. A. Krupa, D. S. Kullakanda, C. X. Shen, and M. R. Warden. Intense threat switches dorsal raphe serotonin neurons to a paradoxical operational mode. *Science*, 363 (6426):538–542, 2019.

[165] S. R. Sesack, A. Y. Deutch, R. H. Roth, and B. S. Bunney. Topographical organization of the efferent projections of the medial prefrontal cortex in the rat: an anterograde tract-tracing study with phaseolus vulgaris leucoagglutinin. *J Comp Neurol*, 290(2):213–42, 1989.

[166] H. Shikanai, T. Yoshida, K. Konno, M. Yamasaki, T. Izumi, Y. Ohmura, M. Watanabe, and M. Yoshioka. Distinct neurochemical and functional properties of gad67-containing 5-ht neurons in the rat dorsal raphe nucleus. *J Neurosci*, 32(41):14415–26, 2012.

[167] P. Soubrie. Reconciling the role of central serotonin neurons in human and animal behavior. *Behav Brain Sci*, 9:319–364, 1986.

[168] K. Spoida, O. A. Masseck, E. S. Deneris, and S. Herlitze. Gq/5-ht2c receptor signals activate a local gabaergic inhibitory feedback circuit to modulate serotonergic firing and anxiety in mice. *Proc Natl Acad Sci U S A*, 111(17):6479–84, 2014.

[169] H. W. Steinbusch. Distribution of serotonin-immunoreactivity in the central nervous system of the rat-cell bodies and terminals. *Neuroscience*, 6(4):557–618, 1981.

[170] T. Sumiyoshi, H. Kunugi, and K. Nakagome. Serotonin and dopamine receptors in motivational and cognitive disturbances of schizophrenia. *Frontiers in Neuroscience*, 8:395, 2014.

[171] R. S. Sutton and A. G. Barto. *Reinforcement learning : an introduction*. Adaptive computation and machine learning series. The MIT Press, Cambridge, Massachusetts, second edition. edition, 2018.

[172] L. F. Takase, M. I. Nogueira, M. Baratta, S. T. Bland, L. R. Watkins, S. F. Maier, C. A. Fornal, and B. L. Jacobs. Inescapable shock activates serotonergic neurons in all raphe nuclei of rat. *Behav Brain Res*, 153(1):233–9, 2004.

[173] S. C. Tanaka, K. Doya, G. Okada, K. Ueda, Y. Okamoto, and S. Yamawaki. Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat Neurosci*, 7(8):887–93, 2004.

[174] S. C. Tanaka, N. Schweighofer, S. Asahi, K. Shishida, Y. Okamoto, S. Yamawaki, and K. Doya. Serotonin differentially regulates short- and long-term prediction of rewards in the ventral and dorsal striatum. *PLoS One*, 2(12):e1333, 2007.

[175] A. Toyama, K. Katahira, and H. Ohira. A simple computational algorithm of model-based choice preference. *Cogn Affect Behav Neurosci*, 17(4):764–783, 2017.

[176] A. Toyama, K. Katahira, and H. Ohira. Biases in estimating the balance between model-free and model-based learning systems due to model misspecification. *Journal of Mathematical Psychology*, 91:88–102, 2019.

[177] L. T. Tran-Nguyen, D. A. Baker, K. A. Grote, J. Solano, and J. L. Neisewander. Serotonin depletion attenuates cocaine-seeking behavior in rats. *Psychopharmacology (Berl)*, 146(1):60–6, 1999.

[178] I. Tsutsui-Kimura, H. Takiue, K. Yoshida, M. Xu, R. Yano, H. Ohta, H. Nishida, Y. Bouchekioua, H. Okano, M. Uchigashima, M. Watanabe, N. Takata, M. R. Drew, H. Sano, M. Mimura, and K. F. Tanaka. Dysfunction of ventrolateral striatal dopamine receptor type 2-expressing medium spiny neurons impairs instrumental motivation. *Nature Communications*, 8(1):14304, 2017.

[179] N. C. Tye, B. J. Everitt, and S. D. Iversen. 5-hydroxytryptamine and punishment. *Nature*, 268(5622):741–3, 1977.

[180] L. Valencia-Torres, C. M. Olarte-Sanchez, D. J. Lyons, T. Georgescu, M. Greenwald-Yarnell, J. Myers, M. G., C. M. Bradshaw, and L. K. Heisler. Activation of ventral tegmental area 5-ht2c receptors reduces incentive motivation. *Neuropsychopharmacology*, 42(7):1511–1521, 2017.

[181] E. J. Van Bockstaele, D. M. Cestari, and V. M. Pickel. Synaptic structure and connectivity of serotonin terminals in the ventral tegmental area: potential sites for modulation of mesolimbic dopamine neurons. *Brain Res*, 647(2):307–22, 1994.

[182] V. Varga, B. Kocsis, and T. Sharp. Electrophysiological evidence for convergence of inputs from the medial prefrontal cortex and lateral habenula on single neurons in the dorsal raphe nucleus. *Eur J Neurosci*, 17(2):280–6, 2003.

[183] S. C. Veasey, C. A. Fornal, C. W. Metzler, and B. L. Jacobs. Single-unit responses of serotonergic dorsal raphe neurons to specific motor challenges in freely moving cats. *Neuroscience*, 79(1):161–9, 1997.

[184] D. Verge, G. Daval, A. Patey, H. Gozlan, S. el Mestikawy, and M. Hamon. Presynaptic 5-ht autoreceptors on serotonergic cell bodies and/or dendrites but not terminals are of the 5-ht1a subtype. *Eur J Pharmacol*, 113(3):463–4, 1985.

[185] P. Vertechi, E. Lottem, D. Sarra, B. Godinho, I. Treves, T. Quendera, M. N. Oude Lohuis, and Z. F. Mainen. Inference-based decisions in a hidden state foraging task: Differential contributions of prefrontal cortical areas. *Neuron*, 106 (1):166–176.e6, 2020.

[186] V. Voon, K. Baek, J. Enander, Y. Worbe, L. S. Morris, N. A. Harrison, T. W. Robbins, C. Ruck, and N. Daw. Motivation and value influences in the relative balance of goal-directed and habitual behaviours in obsessive-compulsive disorder. *Transl Psychiatry*, 5:e670, 2015.

[187] V. Voon, J. Joutsa, J. Majuri, K. Baek, C. L. Nord, E. Arponen, S. Forsback, and V. Kaasinen. The neurochemical substrates of habitual and goal-directed control. *Transl Psychiatry*, 10(1):84, 2020.

[188] K. Vucurovic, T. Gallopin, I. Ferezou, A. Rancillac, P. Chameau, J. A. van Hooft, H. Geoffroy, H. Monyer, J. Rossier, and T. Vitalis. Serotonin 3a receptor subtype as an early and protracted marker of cortical interneuron subpopulations. *Cereb Cortex*, 20(10):2333–47, 2010.

[189] D. J. Walther, J. U. Peter, S. Bashammakh, H. Hortnagl, M. Voits, H. Fink, and M. Bader. Synthesis of serotonin by a second tryptophan hydroxylase isoform. *Science*, 299(5603):76, 2003.

[190] A. Y. Wang, K. Miura, and N. Uchida. The dorsomedial striatum encodes net expected return, critical for energizing performance vigor. *Nat Neurosci*, 16(5): 639–47, 2013.

[191] H. L. Wang, S. Zhang, J. Qi, H. Wang, R. Cachope, C. A. Mejias-Aponte, J. A. Gomez, G. E. Mateo-Semidey, G. M. J. Beaudoin, C. A. Paladini, J. F. Cheer, and M. Morales. Dorsal raphe dual serotonin-glutamate neurons drive reward by establishing excitatory synapses on vta mesoaccumbens dopamine neurons. *Cell Rep*, 26(5):1128–1142 e7, 2019.

[192] J. X. Wang, Z. Kurth-Nelson, D. Kumaran, D. Tirumala, H. Soyer, J. Z. Leibo, D. Hassabis, and M. Botvinick. Prefrontal cortex as a meta-reinforcement learning system. *Nat Neurosci*, 21(6):860–868, 2018.

[193] S. J. Ward, T. W. Lefever, C. Jackson, R. J. Tallarida, and E. A. Walker. Effects of a cannabinoid1 receptor antagonist and serotonin2c receptor agonist alone and in combination on motivation for palatable food: a dose-addition analysis study in mice. *J Pharmacol Exp Ther*, 325(2):567–76, 2008.

[194] M. R. Warden, A. Selimbeyoglu, J. J. Mirzabekov, M. Lo, K. R. Thompson, S. Y. Kim, A. Adhikari, K. M. Tye, L. M. Frank, and K. Deisseroth. A prefrontal cortex-brainstem neuronal projection that controls response to behavioural challenge. *Nature*, 492(7429):428–32, 2012.

[195] B. D. Waterhouse, D. Devilbiss, S. Seiple, and R. Markowitz. Sensorimotor-related discharge of simultaneously recorded, single neurons in the dorsal raphe nucleus of the awake, unrestrained rat. *Brain Res*, 1000(1-2):183–91, 2004.

[196] E. T. Weber and R. Andrade. Htr2a gene and 5-ht(2a) receptor expression in the cerebral cortex studied using genetically modified mice. *Front Neurosci*, 4, 2010.

[197] M. Weber, S. Talmon, I. Schulze, C. Boeddinghaus, G. Gross, H. Schoemaker, and K. M. Wicke. Running wheel activity is sensitive to acute treatment with selective inhibitors for either serotonin or norepinephrine reuptake. *Psychopharmacology (Berl)*, 203(4):753–62, 2009.

[198] B. Weissbourd, J. Ren, K. E. DeLoach, C. J. Guenthner, K. Miyamichi, and L. Luo. Presynaptic partners of dorsal raphe serotonergic and gabaergic neurons. *Neuron*, 83(3):645–62, 2014.

[199] J. S. Wiegert, M. Mahn, M. Prigge, Y. Printz, and O. Yizhar. Silencing neurons: Tools, applications, and experimental constraints. *Neuron*, 95(3):504–529, 2017.

[200] C. A. Winstanley, J. W. Dalley, D. E. Theobald, and T. W. Robbins. Fractionating impulsivity: contrasting effects of central 5-ht depletion on different measures of impulsive behavior. *Neuropsychopharmacology*, 29(7):1331–43, 2004.

[201] Y. Worbe, S. Palminteri, G. Savulich, N. D. Daw, E. Fernandez-Egea, T. W. Robbins, and V. Voon. Valence-dependent influence of serotonin depletion on model-based choice strategy. *Mol Psychiatry*, 21(5):624–9, 2016.

[202] D. E. Wright, K. B. Seroogy, K. H. Lundgren, B. M. Davis, and L. Jennes. Comparative localization of serotonin1a, 1c, and 2 receptor subtype mrnas in rat brain. *J Comp Neurol*, 351(3):357–73, 1995.

[203] S. Xu, G. Das, E. Hueske, and S. Tonegawa. Dorsal raphe serotonergic neurons control intertemporal choice under trade-off. *Curr Biol*, 27(20):3111–3119 e3, 2017.

[204] S. Yagishita. Transient and sustained effects of dopamine and serotonin signaling in motivation-related behavior. *Psychiatry Clin Neurosci*, 74(2):91–98, 2020.

[205] S. E. Yohn, S. L. Collins, H. M. Contreras-Mora, E. L. Errante, M. A. Rowland, M. Correa, and J. D. Salamone. Not all antidepressants are created equal: Differential effects of monoamine uptake inhibitors on effort-related choice behavior. *Neuropsychopharmacology*, 41(3):686–94, 2016.

[206] S. E. Yohn, L. Lopez-Cruz, P. H. Hutson, M. Correa, and J. D. Salamone. Effects of lisdexamfetamine and s-citalopram, alone and in combination, on effort-related choice behavior in the rat. *Psychopharmacology (Berl)*, 233(6):949–60, 2016.

[207] K. Yoshida, M. R. Drew, M. Mimura, and K. F. Tanaka. Serotonin-mediated inhibition of ventral hippocampus is required for sustained goal-directed behavior. *Nat Neurosci*, 22(5):770–777, 2019.

[208] Z. Zhang, J. F. Rickard, S. Body, K. Asgari, C. M. Bradshaw, and E. Szabadi. Comparison of the effects of clozapine and 8-hydroxy-2-(di-n-propylamino)tetralin (8-oh-dpat) on progressive ratio schedule performance: evidence against the involvement of 5-ht1a receptors in the behavioural effects of clozapine. *Psychopharmacology (Berl)*, 181(2):381–91, 2005.

[209] W. Zhong, Y. Li, Q. Feng, and M. Luo. Learning and stress shape the reward response patterns of serotonin neurons. *J Neurosci*, 37(37):8863–8875, 2017.

[210] J. Zhou, C. Jia, Q. Feng, J. Bao, and M. Luo. Prospective coding of dorsal raphe reward signals by the orbitofrontal cortex. *J Neurosci*, 35(6):2717–30, 2015.

[211] L. Zhou, M. Z. Liu, Q. Li, J. Deng, D. Mu, and Y. G. Sun. Organization of functional long-range circuits controlling the activity of serotonergic neurons in the dorsal raphe nucleus. *Cell Rep*, 18(12):3018–3032, 2017.