Identification of an Operant Learning Circuit by Whole Brain Functional Imaging in Larval Zebrafish

Identification of an Operant Learning Circuit by Whole Brain Functional Imaging in Larval Zebrafish

A dissertation presented

by

Jennifer Mengbo Li

to

The Department of Molecular and Cellular Biology

In partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of

Biology

Harvard University
Cambridge, Massachusetts

December 2012

Professor Alexander F. Schier **Jennifer Mengbo Li**

Identification of an Operant Learning Circuit by Whole Brain Functional Imaging in Larval Zebrafish

Abstract

When confronted with changing environments, animals can generally adjust their behavior to optimize reward and minimize punishment. The process of modifying one's behavior based on its consequences is referred to as operant or instrumental learning. Operant learning makes specific demands on the animal. The animal must exhibit some flexibility in its behavior, switching from unsuccessful motor responses to potentially successful ones. The animal must represent the consequence of its actions. Finally, the animal must select the correct response based on its past history of reinforcement. Studies in mammalian systems have found competing and complementary circuits in the cortex and striatum that mediate different aspects of this learning process. The larval zebrafish is an ideal system to extend the study of operant learning due to its genetic and optical properties. We have developed a behavioral paradigm and imaging system that have allowed us to comprehensively image neural activity throughout the zebrafish brain during the process of operant conditioning. Our analysis of the neural network activity underlying this learning process reveals several classes of neurons whose activity correlates with learning and decision making. The distribution of these learning-related neurons was highly localized to regions of the habenula and forebrain. We describe, in particular, a lateralized habenula circuit that may encode prediction and relief prediction error.

Table of Contents

List of Figures

Acknowledgements

First and foremost, I would like to acknowledge Alex for taking a risk on a project in which so many of critical components were uncertain. Alex has given us a unique combination of independence and guidance, for which we are extremely grateful. It has been an invaluable experience to observe the ways in which Alex is able to cultivate innovation in his lab while maintaining scientific rigor, as well as to encourage communication while sustaining a lab of diverse scientific interests.

I would like to thank Florian Engert for his boundless enthusiasm. In particular, his support during the early and very painful phase of working out a new learning paradigm in a difficult organism is greatly appreciated.

I am also very grateful to Jeff Lichtman and Nao Uchida for their many insightful comments over the years. The trajectory of the thesis has been unpredictable at times, with so many technical and experimental uncertainties. They have both helped to direct the project toward the most scientifically interesting and achievable direction. In particular, Jeff's advice to forgo the fictive preparation and proceed immediately to calcium imaging with unparalyzed animals contributed greatly to our progress in the last year and a half.

Both the Schier and Engert labs have given us a great deal of technical and intellectual support, which has made our many projects possible.

Drew Robson contributed equally to all aspects of this project. Drew's ridiculously broad, deep, and ever expanding knowledge of engineering, mathematics, computer science, and biology/neuroscience has made all our projects possible.

Chapter One

Introduction

Introduction

There have been few demonstrations of associative learning in larval zebrafish, and no description of operant conditioning. To put our behavioral paradigm and imaging efforts in context, we will take a broad overview of operant learning paradigms in vertebrate systems, the brain regions implicated in learning based mostly on electrophysiological studies, the recent advances in calcium imaging, and the development of learning and memory circuits over time.

Defining operant conditioning

Operant conditioning and classical conditioning are two of the oldest and most well studied associative learning paradigms[1-5]. In classical conditioning, a neutral conditioned stimulus (CS) is paired with an unconditioned stimulus (US), which activates an unconditioned response (UR). In an ideal paradigm, the CS does not elicit a response before learning. As learning progresses, the CS becomes sufficient to drive a conditioned response (CR) that is often similar to the UR. Pavlov first demonstrated classical conditioning in dogs[4]. A bell (CS) was paired with food (US). Before training, food elicits salivation (UR) but the bell does not. After repeated pairing of the bell and food, the sound of the bell alone can then induce salivation (CR). In operant conditioning, the experimenter selects a response to reinforce with punishment or reward. This training generally occurs in a specific context or is associated with an explicit cue. If the response is rewarded, then its rate of occurrence increases. If the response is punished, then its rate of occurrence declines. Thorndike first demonstrated this in an experimental setting by placing a cat in a puzzle box, which can

only be opened by pushing a specific set of levers[5]. The correct sequence of action that opens the box allows the cat to escape and obtain a food reward. After training, the cat quickly and reliably performs the correct action sequence whenever it is introduced into the box.

Procedurally, classical conditioning is easily distinguishable from operant conditioning. In classical conditioning, the experimenter has control over all relevant stimulus events. The rate of occurrence and duration of both the CS and US are predetermined and do not depend on the animal's behavior. In Pavlov's experiment, the timing of the bell and food is consistent from trial to trial. In contrast, in operant conditioning, the experimenter sets the relationship between an animal's behavior and external stimuli. The animal's behavior dictates the occurrence of stimulus events. In Thorndike's experiment, the cat must perform the correct sequence of action in order to escape. The duration of time the cat spends in the puzzle box varies from trial to trial depending on the cat's actions.

With regards to the underlying learning processes, the differences between classical and operant conditioning have been the subject of long-standing debates, and the exact nature of operant learning has been continually subject to revisions and refinements[1-3, 6-7]. We will take a quick overview of the difference between operant and classical conditioning and the existing frameworks from which to view operant conditioning.

Operant and classical conditioning can produce differences in what an animal ultimately learns, and place different constraints on what an animal can be trained to learn[1-2,6]. In the simplest view, classical conditioning teaches animals stimulus-stimulus (S-S) relationships between the CS and US, meaning that the CS takes on some of the

qualities of the US and therefore activates similar behavioral responses. For example, in Pavlov's experiment, the dog learns to associate food with the bell. In contrast, operant conditioning is believed to teach animals stimulus-response (S-R) relationships, in which an animal learns by trial and error which behavioral response is most favorable in response to a given stimulus. If a given response is positively reinforced, then that S-R relationship is strengthened. If the response is negatively reinforced, then the S-R relationship is inhibited. In the Thorndike experiment, the cat initially performs many erroneous actions that do not open the puzzle box (e.g. trying to squeeze its head through the bars) and only sporadically performs the action sequence that allows it to escape. As training progresses, introduction to the puzzle box immediately drives the correct action sequence, while the erroneous actions are suppressed.

Historically, the debate over the boundaries of classical and operant conditioning appears to have been centered on two main questions[1-2,6]. Are there behaviors and reinforcers that are unique to classical or operant conditioning? Does S-R learning play a role in classical conditioning? With regard to the first question, a wide range of behavioral differences have been proposed -- autonomic vs. skeletal, voluntary vs. involuntary, elicited vs. spontaneous, diffuse and emotional vs. precise and adaptive. Though none of these criteria have been shown to consistently distinguish the learning paradigms, the overall trend does suggest that the behaviors that can be modified by operant conditioning are more varied and flexible, whereas the behaviors that lend themselves to classical conditioning are generally more reflexive and specialized[6]. For example, salivation in the Pavlov experiment is a specialized glandular response to food. In contrast, the lever pressing and pulling actions in the Thorndike operant

conditioning paradigm are motor responses that the animal can continuously adjust. With regard to the second question, it appears that S-R learning is not necessary for classical conditioning. In paralyzed animals that cannot perform the UR, an animal can still learn S-S relationships in some classical conditioning assays. The debate does however highlight the fact that many conditioning paradigms may contain both classical and operant conditioning elements. In many conditioning paradigms, an animal learns both S-S and S-R relationships. Recent studies of the neural basis of operant learning have generally framed behavior in terms of S-R learning.

Habit vs. goal directed learning

Since the initial characterization of operant conditioning as a process of S-R learning, an important distinction has been made in mammalian systems between habitual learning and goal directed learning[7-12]. The key difference between these two learning processes is the existence of an intermediate stage between stimulus and response. In goal directed behavior, the stimulus does not directly evoke a response. Instead, the animal evaluates the potential outcomes associated with possible responses before an action is ultimately selected. The most convincing evidence in support of a two stage process of outcome evaluation and action selection comes from experiments demonstrating the effects of outcome devaluation and contingency devaluation[7,10-12]. In outcome devaluation experiments, an animal is first subject to training trials in which an action (e.g. lever pressing) is associated with a reward (e.g. sucrose). After the animal reliably performs the action, the reward is devalued (e.g. sucrose is made noxious). After exposure to the devalued reward alone, the animal now suppresses the learned behavior. Outcome devaluation experiments were the first line of evidence

demonstrating that animals can represent the consequences of their actions. The learned relationships are referred to as response-outcome (R-O) associations.

Outcome devaluation experiments also demonstrate a distinction between the associations that form early during the learning process and associations that form after extensive training. Goal directed behaviors are generally observed early in the learning process. As the number of training trials increases, behavior becomes less sensitive to outcome devaluation. Holland convincingly demonstrated this phenomenon with a series of rat lever press experiments[10-11]. The animals were trained for 2, 5, or 20 sessions. After 2 or 5 sessions, devaluing the reward with an aversive chemical (LiCl) decreased lever pressing, but after 20 sessions, outcome devaluation no longer had any effect on the rate of lever pressing. This change in sensitivity is thought to be the result of a transition from goal directed behavior to habitual behavior. Unlike in goal directed behavior, habitual responses involve direct S-R associations. The animal no longer evaluates potential outcomes of possible behavior, but rather the stimulus directly elicits the response. In the Holland experiment, after 2-5 sessions, the rat evaluates whether the food reward is desirable before pressing the lever, but after 20 sessions, the presence of the lever directly drives the lever pressing action regardless of the desirability of the outcome.

Evidence for a distinction between goal directed behavior and habitual behavior is mostly restricted to mammalian systems. There is little evidence for this distinction outside of mammalian and avian systems[13]. In fish, operant learning has not been well studied and no conclusive statements can be made based on the paucity of behavioral data.

Motor exploration and switching

In addition to elucidating the differences between goal directed learning vs. habit learning, recent studies have begun to explore a distinct aspect of operant learning – motor exploration and switching[14-16]. While the UR and CR in classical conditioning are constrained by the US selected by the experimenter, operant conditioning paradigms may require animals to explore a range of behaviors to determine their respective R-O relationships. For example, in the Thorndike experiments, the cat initially makes many erroneous movements. As the cat learns the outcomes associated with correct and error movements, the correct movement increases in frequency and the error movements are suppressed.

In some operant conditioning paradigms, the animal is first rewarded for one behavior, and then rewarded for an alternate behavior during a reversal training block. For example, a monkey learns to associate left eye saccades with reward during the first training block. At the beginning of the second training block, the experimenter reverses the reward contingency such that the right saccade is now associated with reward. The monkey is unaware of this reversal of the reward contingency, and will inevitably first saccade to the left during the first trial of the reversal block. After the left saccade is unsuccessful, the animal can either continue to obstinately perform the behavior out of habit or switch its motor response to right saccades. In other words, the animal must overcome habitual responses and switch to a novel behavior. In primates there is some evidence that the anterior cingulate cortex is involved in switching behavior[14-16]. This region is activated during block reversal, and its inactivation disrupts the type of behavioral switching described above.

Brain regions implicated in operant learning

The neural basis of operant learning has been most extensively studied in mammalian systems. Based predominantly on electrophysiological and lesion studies, corticostriatal circuitry mediates many important aspects of the learning process. Electrophysiological recordings in primates and rodents have found signals that correlate with many different aspects of the learning process, such as action selection, outcome evaluation, and reward prediction error (RPE)[17-31]. Action selection and outcome evaluation are believed to be mediated by corticostriatal circuitry[23-31], while RPE is believed to be encoded by the midbrain dopaminergic system[17-22]. Recent studies have also begun to demonstrate the existence of competing corticostriatal circuitry to mediate goal directed vs. habit learning[32-35]. For example, Yin et al. has shown that lesion of the dorsolateral striatum specifically disrupts habit learning but preserves goal directed learning[32]. In fact, if the animal has been previously trained to habitually perform an action, and the dorsolateral striatum is subsequently inactivated, then the animal reverts back to goal directed behavior. This suggests habit learning is mediated by a distinct circuit that competitively suppresses the goal directed learning circuitry.

Though there is a general understanding of the brain regions and cell types involved in operant learning, the exact role of individual regions and neuronal types and their connections to one another is still far from clear. New roles are actively being discovered among cell types that were previously thought to be relatively homogenous. For example, the midbrain dopaminergic system was once thought to predominantly relay reward prediction error signals[17-22]. The view of the dopaminergic system has grown considerably more complex due to a number studies that have found alternative

and more complex roles for dopaminergic neurons[36-46]. Most notably, subsets of dopaminergic neurons have been found to be activated by aversive rather than appetitive stimuli[39-40]. Others have been found to represent stimulus saliency, novelty, and uncertainty[36,38,43-44]. Furthermore, dopaminergic signals can be modulated by past outcomes at different timescales[46]. For example, in a primate reward-biased saccade task, dopaminergic signals can correlate with either the outcome of the immediately preceding trial or with the outcomes of three or more preceding trials.

New roles and connections are also being discovered for individual brain regions. Recent studies of the habenula have suggested a key role in learning[47,59-67]. The habenula is an epithalamic structure that is conserved across many vertebrate species[48]. Structurally, the habenula can be subdivided into lateral and medial habenula, and both parts show significant asymmetry across the left-right axis[48]. The habenula receives inputs from the limbic system and the basal ganglia, and sends projections to inhibit midbrain dopaminergic neurons and activate midbrain serotonergic neurons[48-58]. Interactions with the dopaminergic systems have strongly implicated the habenula in learning[47,58-65]. The lateral habenula in particular has been found to contain neurons that are activated by negative outcomes, and inhibited by positive outcomes[47,49-62]. For example, in monkey eye saccade experiments, the lateral habenula is activated by stimuli that predict a small reward and inhibited by stimuli that predict a large reward[60-61]. Inhibitory projections from the habenula to dopaminergic neurons then shape the response properties of dopaminergic neurons. The discovery of error signals in the habenula and its interaction with the dopaminergic system has

significantly advanced our understanding of the circuitry underlying reinforcement during learning.

Learning systems in teleost fish

The brain regions that contribute to learning are much less understood in fish than in mammals. One key difference in brain development has made it difficult to map brain regions from mammals onto teleost fish. The fish brain undergoes eversion during development, whereas the mammalian brain undergoes evagination. Consequently, brain regions such as the hippocampus, amygdala, and striatum occupy very different positions in the teleost brain[68-70].

Instead of relying on structural similarity, studies using transcription factor expression and lesion experiments have been more successful in coarsely defining some correspondence between fish and mammalian brain regions. Based on gene expression in adult teleost animals, it is believed that the dorsal lateral pallium is homologous to the hippocampus, the dorsal medial pallium is homologous to the amygdala, and the subpallium is homologous to the striatum[68-70]. Other regions are more easily defined, such as the habenula, which is structurally similar between fish and mammals[48,71-72]. Gene expression indicates that the fish habenula can also be divided into medial and lateral habenula, though the relative position of these cells within the habenula differs from mammals[72]. Key catecholinergic systems are also beginning to be mapped in greater detail in zebrafish, and structures such as the midbrain dopaminergic system, raphe nucleus, and locus coeruleus have been described[73-77]. The problem of homology becomes even more difficult to address at larval stages. In larvae, the location of most

telencephalic structures such as the hippocampus, amygdala, and striatum are not well understood. Outside of the telecephalon, some prominent structures can be identified by anatomy even in larval zebrafish, such as the habenula in the epithalamus and the cerebellum in the rhombencephalon[71-72,78]. Catecholinergic centers such as the raphe and midbrain dopaminergic cluster can also be identified by gene expression, but they are not easily identifiable by anatomy alone[73-75].

Beyond establishing basic homology using gene expression, the teleost field has employed lesion experiments in an attempt to demonstrate that different brain regions underlie different learning processes. In goldfish, lesions in the lateral pallium (homologous to the hippocampus) impair spatial learning and trace conditioning, but leave classical conditioning and fear conditioning intact[80-81]. This is consistent with the evidence from mammals that the hippocampus encodes relational memory. In contrast, lesions of the goldfish medial pallium impair avoidance learning but do not affect spatial learning[82-83]. Lesioning the cerebellum appears to broadly affect learning, causing defects in classical conditioning, spatial learning, and fear conditioning[84-86]. More recently, the habenula has been implicated in anxiety-mediated behavior and learning[72,87]. For example, Agetsuma et al recently inactivated the dorsolateral habenula in adult zebrafish. The ablated animals show a bias towards freezing rather than escape during avoidance learning[87]. This suggests that the dorsolateral habenula is involved in the selection of fear or stress responses in zebrafish.

Overall, these studies represent fairly crude efforts to isolate the neural circuitry that underlies learning. Lesion studies disrupt entire brain regions, despite the fact that each region contains a mixed population of neurons. Lesion studies cannot assess how these

neurons interact as a network. There has yet to be any conclusive evidence for error signals, salience signals, or action selection signals in teleost systems. Studies of learning in fish have also been limited in several ways. One, compared to mammals, relatively few learning paradigms have been developed for fish. In adult zebrafish, visual association tasks and spatial learning have been demonstrated in a plus maze paradigm[89], and operant learning has been shown in one study using a shuttle box[96]. Two, there have been no successful applications of tetrode recordings to fish during learning. Three, while the larval zebrafish has been shown to be a particularly amenable system for monitoring network activity during behavior, learning assays have been particularly difficult to develop for this system[95-96]. In the next chapter, we will introduce an operant learning paradigm in larval zebrafish that takes advantage of the potential for whole brain calcium imaging in this system.

Calcium imaging

Recent advances in genetically encoded calcium indicators have allowed experimenters to record an order of magnitude more neurons at single cell resolution than was possible with electrophysiological techniques[97-101]. Though still somewhat limited in temporal resolution, calcium imaging is a non-invasive technique that gives us an unprecedentedly comprehensive view of neural networks.

Application of calcium imaging to studies of behavior requires the development of techniques and assays that permit animals to behave while physically restrained under a microscope. Recently, this has been successfully implemented in several model systems. Most notably, brain regions such as the hippocampus, parietal and motor

cortex have been successfully imaged in head-restrained mice during spatial and motor learning[102-106]. These groundbreaking studies are based on a combination of virtual reality, two photon imaging, and GCaMP3. In Drosophila, techniques have been developed to image the brains of head-fixed flies while they walk on an air-supported ball[106]. Head-restrained flies in a flight simulator have been successfully trained using both classical and operant conditioning paradigms[107-108]. In zebrafish, fictive motor recordings in paralyzed animals have been used to image whole brain activity during motor adaptation[109]. In this preparation, recordings from the motor nerves of paralyzed larval zebrafish are used to decode motor activity. This setup was combined with calcium imaging to uncover signals related to motor adaption. While this setup minimizes motion artifacts, it is limited by an incomplete readout of motor activity as well as by drug-induced suppression of acetycholinergic activity throughout the brain.

We have developed an operant learning paradigm in awake behaving larval zebrafish that is amenable to simultaneous calcium imaging. After generating a transgenic line that expresses GCaMP5 under a pan-neuronal promoter, we imaged neural activity across the entire brain during learning. We were then able to ask several fundamental questions about the circuit activity that mediates operant learning in zebrafish. 1) Can we identify neuronal signals that correlate with learning? In mammals, prediction and prediction error signals have been identified for both reward punishment. These signals are not sensitive to reward contingency, but are highly sensitive to the past performance of the animal (i.e. whether the previous trials were correct or incorrect). 2) Can we identify signals that correlated with decision making? Unlike prediction related signals, signals encoding action selection or R-O associations should show sensitivity to reward

contingency, and discriminate between alternate courses of actions. 3) How are learning-related signals distributed throughout the brain? Are they homogeneously distributed across the entire brain, or are there discrete learning centers? Are all learning signals bilaterally symmetric or are there left-right differences in the distribution of learning-related signals? 4) Does the presence or absence of specific functional clusters correlate with learning ability? Specifically, we ask whether we can isolate the class of neural activity that is absent or disrupted in animals with impairments in learning. 5) How does the global network activity change during learning? Is there a continual shift in activity across successive trials, or are there abrupt shifts in activity during learning? Can we identify the neurons that underlie the changes in network activity?

Chapter Two

Operant learning in larval zebrafish

Abstract

When confronted with changing environments, animals generally adjust their behavior to optimize reward and minimize punishment. The process of modifying one's behavior based on observed consequences is referred to as operant or instrumental learning. Work in mammalian systems has uncovered some fundamental rules that govern the process of operant learning, and has begun to elucidate the underlying neural circuits that govern different aspects of this complex learning process. Larval zebrafish, with its genetic and optical properties, is an ideal system to extend and complement the work on operant learning in mammalian systems. However, demonstration of learning has been sparse in larval zebrafish, and demonstration of operant learning non-existent. We have developed a simple operant learning paradigm in larval zebrafish that is amenable to simultaneous calcium imaging. We show that this behavior is rapidly acquired, stably retained, and readily reversible. Furthermore, we characterize the emergence of learning ability across larval development, as well as two naturally occurring failure modes of the learning circuit.

Introduction

Operant learning is a process by which an animal modifies its behavior based on the consequences of its actions[1-2,5-7]. Unlike typical forms of classical conditioning, the response is not reflexively driven by an external stimulus. Instead, the animal must go through a process of trial and error until the correct response is discovered.

In mammalian systems, operant learning proceeds through two distinct phases[6-7]. In the early trials of a training block, animals are often required to explore multiple behaviors.

This often requires overcoming innate or habitual responses. The animal must represent the outcome of possible responses, as well as the value of each outcome[7-12]. If an outcome is favorable, then the R-O association that led to the outcome is reinforced. In this initial phase, if an outcome becomes devalued, the response associated with that outcome is immediately suppressed. As training progresses, the correct response becomes habitual[10]. At this stage, the animal no longer evaluates the outcomes of possible responses. Instead, external sensory cues directly drive the correct response.

In zebrafish, learning has not been well studied. In adult zebrafish, the learning paradigms that have been developed have mainly centered on fear learning, spatial learning, or classical conditioning[89-96]. At the larval stages, there have been virtually no demonstrations of learning, with the exception of one study on classical conditioning[95]. The authors use a classical conditioning approach using a visual stimulus as the CS and touch as the US. Ten paired CS + US trials separated by 6 min were followed by three test trials after various delays. The authors demonstrate that the pairing protocol potentiates behavioral responses to the CS 6-fold, while unpaired controls were relatively less potentiated. The effect is strongest within 30 min of the pairing and becomes significantly reduced by 60 min. Using calcium imaging with synthetic dyes, the authors claim that the pairing protocol non-specifically increases cerebellar activity. Optical ablations of the entire brain region also show deficits in the behavior. While we have been unable to replicate the behavioral results of this study, we believe the main limitation in any case is that it treats an entire brain region as a homogeneous structure instead of attempting to identify distinct functional types. The advantage of establishing

a zebrafish learning paradigm at the larval stage is the unprecedented access to individual neurons throughout the circuit. To advance our understanding of learning beyond the adult literature, the nascent larval learning field needs to identify distinct functional roles played by individual neurons in the context of a complete learning circuit.

To take advantage of the expanding array of tools of imaging and manipulate neural activity in larval zebrafish, we have developed an operant learning paradigm that is amenable to simultaneous calcium imaging throughout the brain. We have demonstrated that larval zebrafish are capable of rapid acquisition and stable retention of the learned response. We have shown that larval zebrafish can also overcome this learned response and switch their motor responses when the R-O contingency is reversed. Finally, we have traced the emergence of this learning ability across larval development.

Operant Learning Paradigm

In order to eventually image neural activity throughout the brain during learning, we sought a learning task that could be performed by head-restrained animals under a microscope. We have succeeded in conditioning head-restrained animals to bias their turn direction in response to noxious heat. In this learning task, we tether a larval zebrafish by embedding its head in agarose, while its tail is allowed to move freely. Tail movements are tracked at 200 frames per second and then decoded into turns and swims. An IR laser is aimed at the head and body of the animal to deliver heat pulses at 32 °C.

The learning task is divided into two blocks, with 25-30 trials in each block (**Fig. 1a**). At the start of each trial, the laser is turned on, which typically elicits turns within 5 s. If the first turn is in the correct direction, the laser is turned off and the trial is scored as a correct trial. If the first turn is in the incorrect direction, the trial is scored as an error trial and the laser stays on until a subsequent turn is in the correct direction. All training trials in the first block have the same R-O contingency, which is chosen to oppose any preexisting turn direction preference the animal might have (see methods). In the second block, the correct turn direction is reversed (e.g. if the first block rewards left turns, then the second block rewards right turns). In order to minimize noxious heat exposure, animals must learn to bias their turn direction in response to heat onset, and must learn to reverse this bias during the second block.

Turn direction bias is rapidly acquired

We find that larval zebrafish rapidly learn to adjust their turn direction in response to noxious heat (**Fig. 1a-b**). Several performance metrics demonstrate that learning is occurring over time. 1) The most direct measure of learning is the correctness of the first turn direction after heat onset. We score each trial as correct if the first turn is in the direction of the heat off contingency. This is a stringent metric because trials are only scored as correct based on the first motor event, even if the animal is able to subsequently turn off the heat after the initial error turn. Across the entire population of animals, the percentage of correct trials increases rapidly during training. On average, only 6 trials are necessary to entrain a significant bias in the correct turn direction (**Fig. 1c**). 2) Another performance metric is the duration of heat exposure in each trial. If an animal successfully learned to bias turns in the correct direction, then the animal would

succeed in reducing the duration of the laser exposure in successive trials. We find that the duration of laser exposure decreases quickly across trials (**Fig. 1d**). The largest reduction in exposure time occurs after the first training trial. 3) We can define another learning measure based on primate learning paradigms, where it has been shown that higher certainty can lead to reduced reaction times after stimulus presentation. We measured the time of the first turn after heat onset, and find a negative correlation of response latency to past performance (defined as the fraction of the past five trials that were correct) (**Fig. 1e**). These three metrics collectively indicate that larval zebrafish can rapidly learn to adjust their turn direction in response to noxious heat.

The learned turn direction bias is also readily reversible. The percentage of correct trials reaches a similar plateau in both training blocks (**Fig. 1c**). This result indicates that, in the reversal block, animals can override the learned response, and switch to a new motor program.

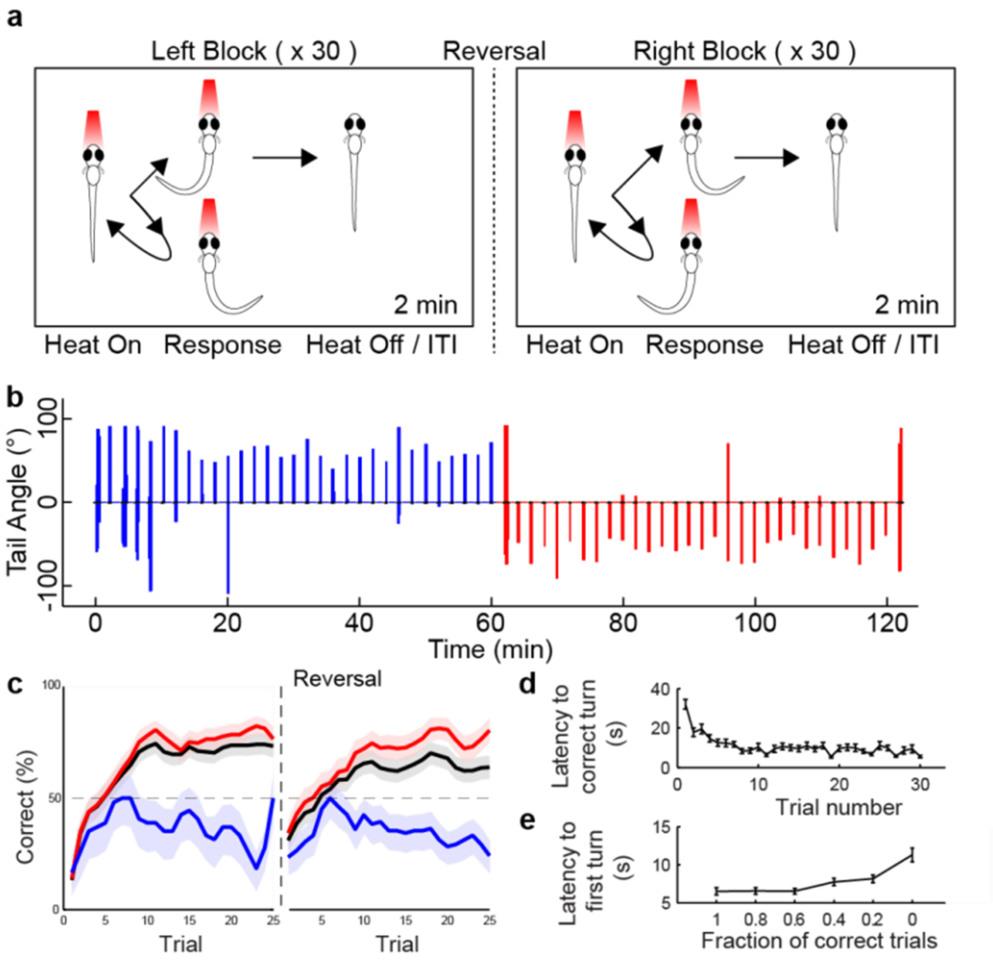

Figure 1 | An operant learning paradigm for larval zebrafish. a, Schematic of the learning paradigm. In each trial, larvae are heated with an invisible infrared laser until they perform a tail flick in the rewarded direction. After a block of 25-30 trials with the same reward contingency, the rewarded direction is reversed without notice and another 25-30 trials are conducted with the reversed reward contingency. The rewarded direction in the first block is chosen to oppose to any innate turn direction bias in a given animal (see Methods). **b,** Example of a 7 dpf larval zebrafish performing operant learning in the paradigm. In early trials of the first block, the first tail flick is incorrect but a correct tail flick occurs later in the trial and terminates the heat. After five error trials, the first tail flick is correct for the most of the remaining trials of the block. In the first trial of the reversal block, the first tail flick is in the rewarded direction of the first block, which is now incorrect. A correct tail flick later in the trial terminates the heat. In the remaining trials, the first tail flick is almost always correct. **c,** Learning performance averaged across all 7-8 dpf zebrafish (black lines, n = 191), as well as the subpopulations we classify as learners (red lines, n = 134) and non-learners (blue lines, n = 57, see Methods). A trial is scored as correct if the first tail flick was in the rewarded direction. **d,** Latency to the first correct turn decreases across trials within each block. **e,** Latency to the first turn depends on past performance, defined as the fraction of correct trials in the last five trials. All error bars represent s.e.m.

Turn direction bias is retained for one hour and is stimulus dependent

To test whether larval zebrafish can retain the learned turn direction bias, we trained animals for one block (30 trials), followed by a rest period of one hour (**Fig. 2a**). We then determined the percentage of correct trials immediately following the rest period across a population of trained animals. If the animal retained the learned turn direction bias, then performance in the first trial following the rest period should be significantly above chance. We find that the percentage of correct trials following the rest period is >85%, indicating that the animal retained the bias that it learned during the training block (**Fig. 2b**).

The rest period also allows us to test whether the turn direction bias we observed in response to heat onset is stimulus dependent. There are two possible outcomes that can be distinguished. One possibility is that animals simply learn to bias all turns toward the correct direction regardless of context. Alternatively, animals could learn to bias turns toward the correct direction only in response to noxious heat. In the first case, we would expect all turns, spontaneous or heat driven, to be biased toward the correct turn direction. In the second case, we would only expect heat driven but not spontaneous turns to be biased. We find that spontaneous turns generated during the rest period are not biased toward the previously trained turn direction (**Fig. 2c**), indicating that the learned response is context dependent. This result establishes that our learning assay falls under the category of S-R learning as expected for an operant conditioning paradigm (reviewed in Chapter 1).

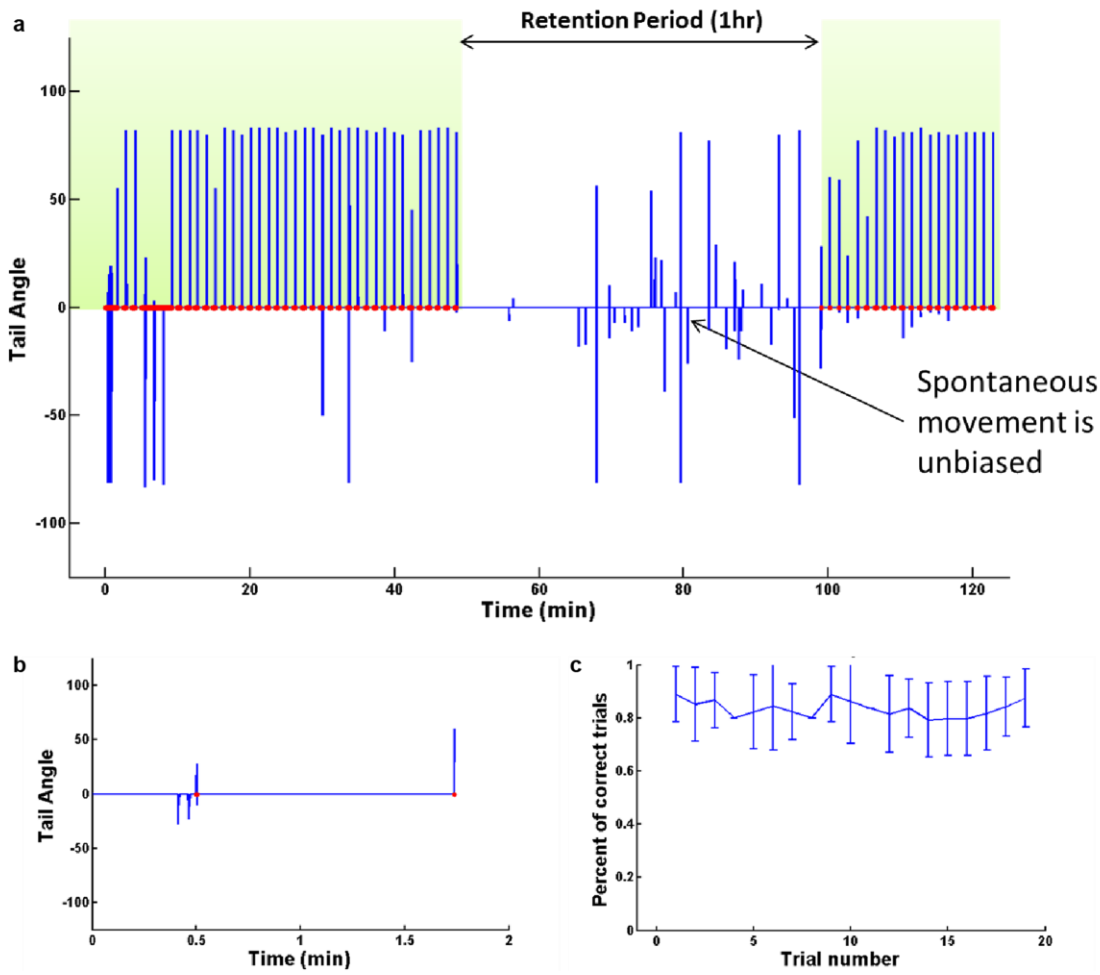

Figure 2 | Learned response is retained for 1 hour and is stimulus dependent

a, Behavioral response during a retention protocol consisting of a block of 30 training trials, followed by a 1 h retention period, after which we administer test trials with the same R-O contingency. **b,** Initial tail movements in first test trial after retention block. **c,** Average performance during test trials after retention block.

A subset of animals show distinct types of learning impairment

The learning curve indicates that performance after 25 training trials is not one hundred percent correct. The learning curve plateaus at around 75%. This could arise in multiple ways. At one extreme, it could be that all animals learn, but learn imperfectly, peaking at 75% correct. At the other extreme, a subset of animals might perform perfectly after

training, while smaller subset of animals do not exhibit any learning after training. Alternatively, an intermediate explanation is also possible, in which a subset of animals learn perfectly, a subset learn imperfectly, and a subset do not learn at all.

To distinguish between these possibilities, we fit the learning performance of the animals to a sigmoidal curve. This allows us to determine the asymptotic performance of the animal. An animal that learns perfectly would have an asymptotic performance index of 1. An animal that fails to learn would have an asymptotic performance index ≤ 0. We find that the distribution of asymptotic performance index across the population is bimodal (**Fig. 3a**). A subset of animals show clear learning, with an average asymptotic performance index of around 0.85, while another subset of animals show no learning, with an average asymptotic performance index of around 0.5.

We further examined the subset of non-learners, and found that they fall into two distinct categories based on their behavior during error trials. In an error trial, animals make a series of error turns, before eventually switching their turn direction and terminating the heat exposure. We find that the duration of heat exposure in error trials also exhibits a bimodal distribution (**Fig. 3b**). Most animals turn off the laser within one minute, with a mean duration around 15s. However, a subset of animals fails to switch from error turns to correct turns for an extended period of time, resulting in a mean duration around 110s. We propose two different types of non-learners with distinct behavioral defects.

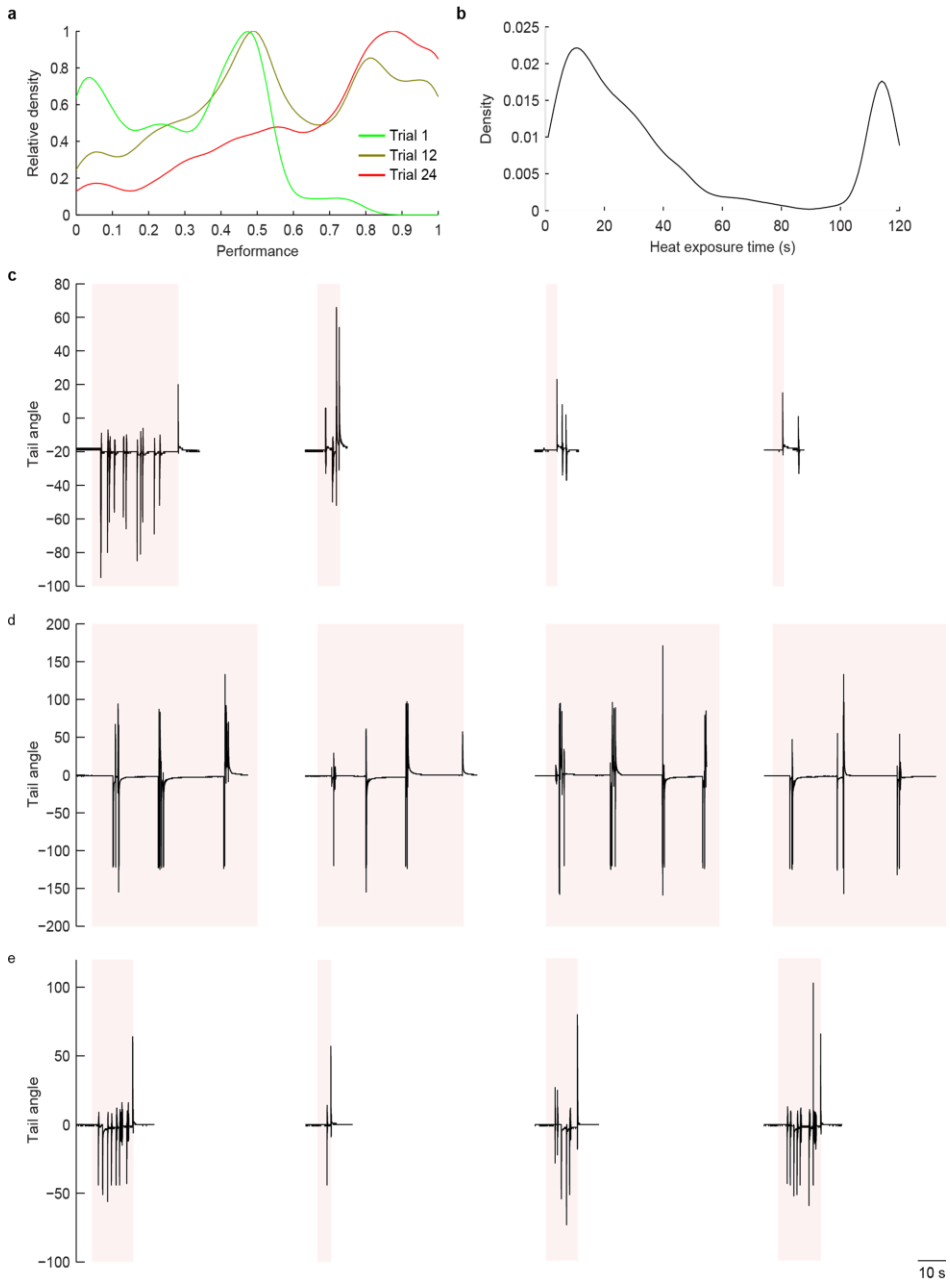

Figure 3 | Switching-impaired and retention-impaired animals

a, Relative distributions of performance across fish in early (trial 1), middle (trial 12) and late (trial 24) trials of a training block. Mode near 0 reflects innate turn bias that we train against in the first block (see Methods). Mode near 0.5 reflects chance performance. Mode near 1 indicates learning. With increasing trial number, learners transition through all three modes. Note that by trial 24, a subset of fish remain distributed around 0.5, which we define as retention-impaired animals. **b,** Distribution of exposure times to noxious heat in error trials. We define animals as switching-impaired if their average exposure time exceeds 60 s. **c-e**, typical examples of four sequential trials in a learner (**c**), a retention-impaired animal (**d**), and a switching impaired animal (**e**).

Switching-impaired animals

After an animal generates one or more error turns near the beginning of an error trial, the animal typically switches to a turn in the correct direction within 20s. However, some animals fail to switch their turn direction for over 1 minute in error trials (**Fig. 3b,d**). We define this group of poor learners as switching-impaired animals. This difference is not likely to be the result of a sensory deficit (i.e. failure to detect laser) because the latency of the first movement after heat onset is not significantly different from that of learners. Instead, we propose that a switching signal is lacking in the learning circuit of this class of animals. We note that the asymptotic performance of these animals can be as good as learners. This result suggests that switching-impaired animals do not necessarily have impaired retention of R-O relationships once they ultimately try the correct turn direction. To avoid this ambiguity, switching-impaired non-learners were excluded from our comparison of neural activity between learners and non-learners (see chapter 3).

Retention-impaired animals

Some animals never learn the task within the finite number of trials in our assay. We refer to this group as retention-impaired animals, defined by asymptotic performance ≤ 0.5. This phenomenon occurs despite the fact that some of these animals show robust switching between turn directions during error trials (**Fig. 3a,e**). We propose that a retention signal is lacking in this class of animals.

We note that our definition of these two impairments are not mutually exclusive. Our findings suggest that there are separable signals in the learning circuit such as

switching signals and retention signals. We will explore both types of signals in the next chapter.

Ability to learn develops over time

Our learning assays are generally performed at 6-7 days post fertilization (dpf). To determine when larval zebrafish become capable of performing this learning task, we assayed learning performance from 3 dpf to 7 dpf. We find that asymptotic learning performance improves dramatically from 3 dpf to 6 dpf, with little change between 6 dpf and 7 dpf (**Fig. 4**). We find both types of learning impairments from 3-7 dpf, though a greater proportion of these animals tend to be retention-impaired from 3-5 dpf. Our results suggest that the ability of larval zebrafish to learn matures quickly over the course of 3 days from 3dpf to 6dpf.

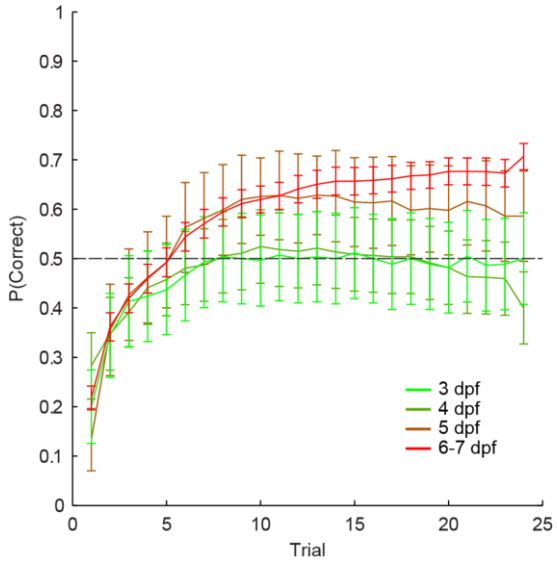

Figure 4 | Learning curves across larval development

Larval learning at 3dpf, 4dpf, 5dpf, and 6-7 dpf. Performance metric is probability that the first turn of a given trial is in the correct direction.

Open loop control

In an operant learning task, correlation between behavioral output and reward/punishment is essential to learning. To test whether this principle is true in our task, we used an open loop assay to demonstrate that learning only occurs when the behavioral output is correlated with reward/punishment. In the open loop assay, we randomized the duration of heat exposure for each trial. Consequently, heat offset is no longer contingent on the animal making the correct turn.

After 25 trials in an open loop block, animals either retain the turn direction bias from before the trials begin, or the turn direction becomes random from trial to trial (**Fig. 5a**). Since no turn direction is rewarded, we cannot score trials as correct or incorrect. Nevertheless our results demonstrate that exposure to heat alone cannot change the turn direction bias of an animal, but that feedback is necessary to entrain an animal to make the correct turn in response to noxious heat.

Our open loop results also demonstrate that the reaction time of the animal to heat onset becomes longer (**Fig. 5b**). This is consistent with our observation that as learning progresses, animals respond more quickly to heat onset. In the open loop block, as number of trials increases, the relationship between motor movement and sensory feedback weakens. There is increased uncertainty regarding the effect of each movement, and reaction time slows.

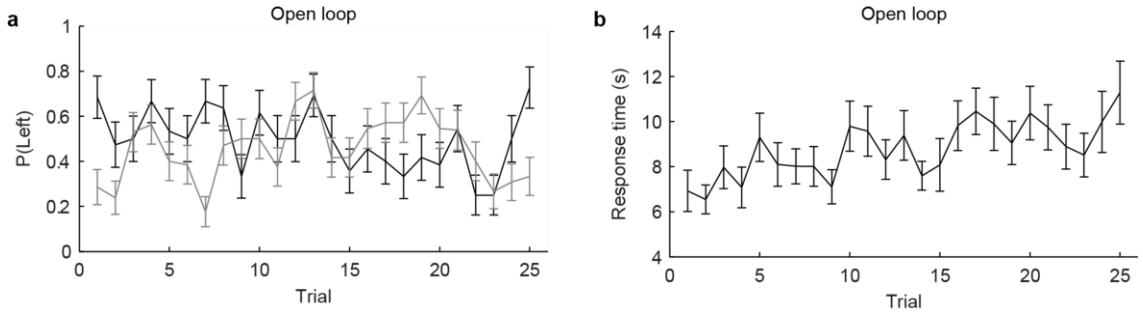

Figure 5 | Open loop control after block training

a, Probability of left turn during 25 trials of an open loop protocol performed immediately after the second block of operant conditioning, if the R-O contingency of the second block rewarded left turns (black) or right turns (gray). In each trial, the duration of heat exposure was randomly sampled from a uniform distribution between 0 and 1 min, effectively uncoupling thermal experience from the motor output of the animal. **b,** Time until first motor event after start of each open loop trial.

Conclusion and discussion

Our results demonstrate that larval zebrafish are capable of modifying their behavior based on the consequences of their actions. More specifically, we show that zebrafish can learn to bias their turn direction in response to noxious heat to minimize heat exposure. We find this learned response to be rapidly acquired and stably retained for at least one hour. Using a reversal training block, we also show that larval zebrafish can overcome a learned response to switch to a new behavioral response during reversal training. Finally, we find that the ability to learn this task emerges between 3-6 dpf.

Our results suggest that the ability of larval zebrafish to learn matures quickly over the course of 3 days from 3-6 dpf. This raises several questions that we will explore in Chapter 4. Are there circuit elements that emerge between 3-6 dpf which are correlated with learning performance? For example, is there a retention signal that is absent at 3 dpf but present at 6 dpf? Is there a difference in the maturation rate of different brain

regions with respect to learning? Finally, is there any similarity in neural activity between learning impairments at 7 dpf and learning impairments at 3 dpf? In other words, can the failure to learn at 7 dpf be attributed to delayed maturation of certain circuit elements?

Methods

Head-restrained behavior

Zebrafish larvae between 3-8 dpf were embedded in 3% intermediate melting point agarose and then freed caudal to the swim bladder so that the tail could move freely. After mounting, animals were maintained at room temperature for 10 hours (overnight) and then subjected to experiments the following day. After overnight incubation, animals were inspected for damage due to the mounting procedure.

Thermal stimulus delivery

To rapidly deliver noxious heat to head-restrained animals, we used a 1 W 980 nm fiber-coupled laser (Roithner Lasertechnik) that was collimated to a beam size of roughly 0.5 mm (Aistana). After loading a previously mounted larval zebrafish under the microscope, we manually positioned the laser to illuminate the front of the head from 4 mm in front of the fish while carefully centering the left/right axis to ensure symmetry across the midline. Thermal waveforms were generated in real time using a data acquisition board (National Instruments) and a 2 A laser driver (Thorlabs).

Real-time tail tracking

To monitor motor output, we illuminated from the side with a 950 nm LED while imaging at 200 Hz (Point Grey). To avoid cross-talk from the thermal stimulus laser, we used a 950 nm short pass filter at the camera entrance. Starting from a manually defined anchor point on the trunk of the animal, the software performed a series of line searches at a fixed segment radius to obtain a skeletal representation of the midline of the tail. We defined the tail angle as the angle made by the tip of the tail with respect to the anchor point, defining tail straight as 0°.

Motor decoding

To decode tail movements into swims and turns, we employed a virtual trajectory simulator that we recently developed to study head-restrained navigation (Robson and Li, unpublished). The simulator uses a finite state machine that detects extrema in the tail angle in real time. When a new extremum θ_i is identified at time t_i, the simulator estimates of the power $\bar{\theta}_i = (\theta_i + \theta_{i-1})/2$ and asymmetry $\Delta\theta_i = \theta_i - \theta_{i-1}$ of the tail movement from θ_{i-1} to θ_i. After processing each camera frame, the simulator updates a model of the virtual linear velocity according to

$$\frac{dv}{dt} = \sum_i f(\bar{\theta}_i, \Delta\theta_i)\delta(t - t_i) - cv$$

and the virtual heading according to

$$\frac{d\theta}{dt} = \sum_i g(\bar{\theta}_i, \Delta\theta_i)\delta(t - t_i) - c\theta$$

Transfer functions f, g and decay rate c were manually determined based on free swimming data.

Forward swim events are characterized by very low asymmetry and predominantly drive linear velocity. Turn events consisting of individual tail flicks predominantly drive angular velocity. Turn events that are followed by counterturns or forward swim movements contribute to angular velocity and linear velocity. In this study, we detected turn events when the angular velocity exceeded a threshold. We scored only the initial turn direction, after which the detector was disabled until the tail was inactive for at least 2 s.

Operant conditioning protocol

Operant conditioning was performed in blocks of 25-30 trials. Trials had a fixed length of 2 min. At the start of each trial, we turned on the thermal stimulus laser. We used a laser power of 150 mW, which was calibrated to raise the temperature of the animal to approximately 32 °C. In each block, we monitored the motor output of the animal and turned the heat off the first time that the animal initiated a turn to the direction of the reward contingency. When the animal performed multiple tail deflections in close succession (<100 ms), we specifically scored the direction of the first deflection to prevent the animal from spuriously inactivating the laser by counterturns or struggles. Prior to the first block, we performed a single probe trial to determine the naïve turn direction preference of each animal. We set the reward contingency direction of the first block to be against this naïve turn direction preference. The entire training protocol took approximately 2 hours.

Chapter Three

Circuit dynamics and functional classes that underlie operant learning in larval zebrafish

Abstract

Operant learning is a process by which an animal updates its predictions about the external world and biases its actions to optimize reward and minimize punishment. In mammalian systems, the basal ganglia have been firmly established as an essential site of reward prediction and decision making. However, the diversity of learning-related signals and their interactions across brain regions are still being continually refined. By combining operant learning and brain wide imaging in larval zebrafish, we take an unbiased approach to uncovering learning-related activity throughout the entire brain. Using this approach, we were able to identify both a novel habenula network that encodes relief prediction and prediction error, as well as a collection of action selection neurons that implement the learned responses. Activation and suppression of these neurons correlates with learning induced changes in behavior, and the depletion of specific functional classes is correlated with naturally occurring learning deficits. Based on these results, we propose a provisional forebrain circuit for operant learning in a simple vertebrate system.

Introduction

The neural basis of operant learning has been most extensively studied in mammalian systems. One of the most significant developments over the past several decades has been the characterization of two canonical learning-related signals – reward prediction (RP) and reward prediction error (RPE)[17-22]. Reward prediction signals are activated by reward predicting cues, and are positively modulated by past performance, meaning that these signals are higher when the preceding five trials were mostly correct but

lower if the preceding five trials were mostly incorrect. Reward prediction error neurons shift from responding to actual reward to responding to reward predicting cues. The reward triggered activity is negatively modulated by past performance, while the cue-triggered activity is positively modulated by past performance.

Though significant progress has been made toward identifying and characterizing RP and RPE neurons in mammals, a number of important questions remain to be fully resolved. 1) What are the types and properties of prediction and prediction error signals in non-canonical paradigms such as relief from noxious stimulus?[136-140] 2) Are there distinct neuronal populations that encode the presence and absence of relief prediction? 3) To what extent are prediction and prediction error signals conserved in non-mammalian systems? Prediction error signals have been found in Drosophila but remain relatively uncharacterized other non-mammalian systems[141].

We introduce for the first time a learning paradigm in larval zebrafish that allows for simultaneous brain wide imaging of neural activity during operant learning. Using this system, we identify a functionally asymmetric habenula circuit that encodes the predictability of relief. We find three main components to this system: neurons that encode relief prediction, neurons that encode the absence of relief prediction, and neurons that encode relief prediction error. By comparing zebrafish that learn to a subset of animals that fail to learn, we find that disruption of prediction-related signals and action selection signals is correlated with failure to learn new reward contingencies. We propose that the habenula relief prediction circuit acts in concert with action selection neurons to bias turn direction and modulate response latency during operant learning. These results represent the first characterization of relief prediction and action

selection signals in a non-mammalian vertebrate brain, and suggest a possible functional lateralization in the habenula that remains to be explored in other systems.

Single cell analysis reveals relief prediction-related signals

To identify neuronal subpopulations that are active during learning, we have developed an imaging pipeline that automatically identifies task-related neural activity (**Fig 6**). 1) We employ both online and offline image registration to stably monitor neural activity at single cell resolution while tethered animals perform the operant learning task. 2) We record both the thermosensory stimulus and locomotor output in order to update the behavioral paradigm as well as correlate sensory and motor events with neural activity. 3) We automatically identify cell centroids, extract an activity trace for each cell, and perform linear regression to identify heat on, heat off, inter-trial interval (ITI), and motor-related components of each activity trace. From a pooled data set of 208 animals, we identified 16079 task-related neurons in the forebrain, 4519 task-related neurons in the midbrain, and 3948 task-related neurons in the hindbrain.

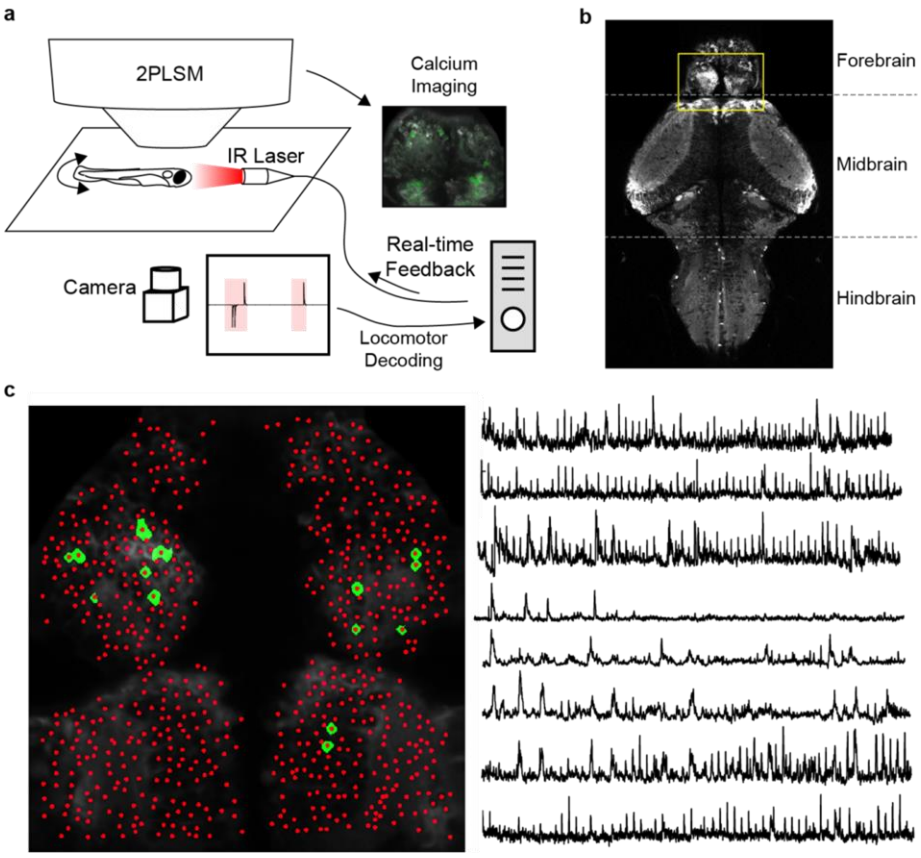

Figure 6 | Brain-wide functional imaging during learning. a, Two photon laser scanning microscopy during closed loop tail-free behavior. **b,** After online and offline image registration, an automated image analysis pipeline generates anatomically-defined cell centroids and activity-defined cellular masks (left). Calcium activity traces are extracted at single cell resolution and provide stable recordings for hours (right).

In order to identify neurons with learning-related activities, we constructed a response vector for each neuron that summarizes its correlation to the learning curve, its correlation to past performance (i.e. fraction of correct trials in the past five trials), and its peak activity in correct and error trials of each block (**Fig. 7**). For heat on-related activity, we find that 33% of neurons show >0.1 correlation to past performance, and 17% show <-0.1 correlation to past performance (**Fig. 8c**). For heat off-related activity, 16 % of neurons show >0.1 correlation to past performance, and 13 % show <-0.1

correlation to past performance (**Fig. 8d**). A majority of neurons (71%) also show differential activity between error and correct trials for both heat on and heat off related activity (**Fig. 8g-h**).

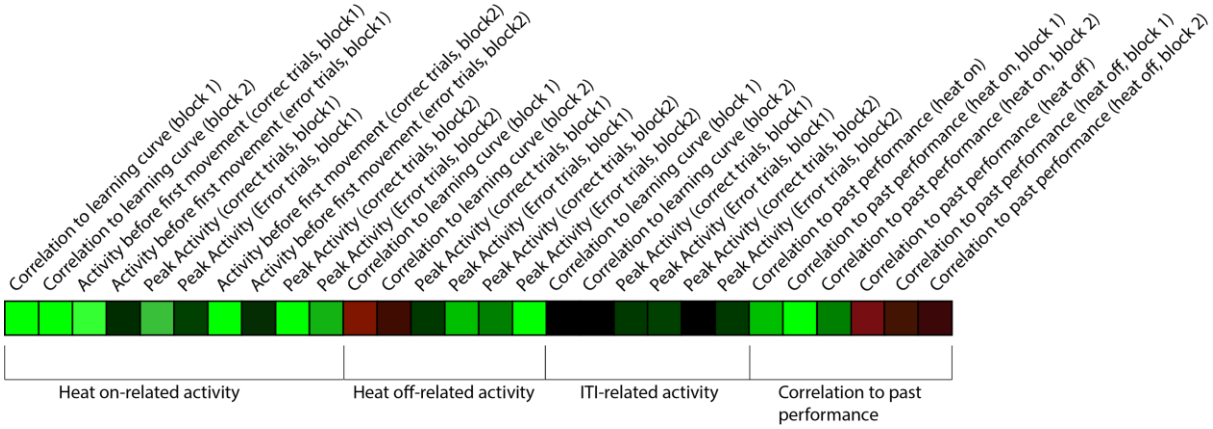

Figure 7 | Schematic of a response vector for a given cell. The response vector provides a summary of the peak activity and correlation metrics for a given neuron.

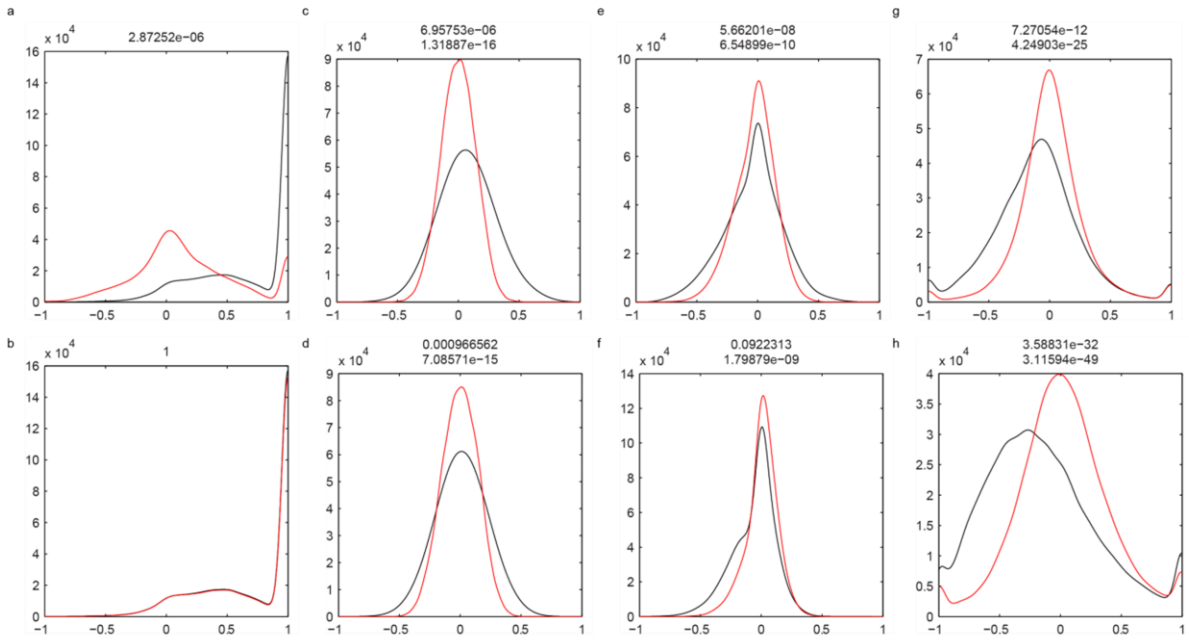

Figure 8 | Distribution of activity and correlation values for scrambled (red) and un-scrambled datasets (black). a, Distribution of ITI and stimulus specific activity for the original dataset (black), and circularly permutated dataset (scramble method 1, red). Each neuron is scored on a specificity index from -1 to 1. A score of -1 means that all activity occurs during the ITI period, a score of one indicates that all activity occurs around a stimulus triggers. Scrambling by circular permutation significantly reduces the number of neurons with stimulus specific activity (score of > 0). P-value is shown at the top (two sample t-test). **b,** The same comparison of activity specificity as in **a** but against a dataset scrambled by trial order (method 2). Scrambling by trial order does not change activity specificity since all relationships with sensory and motor trigger points are maintained. **c-d,** Distribution of correlations to past performance for heat on activity (**c**) and heat off activity (**d**). Original dataset is in black, scrambled dataset (by trial order) is in red. Scrambling by trial order reduces the number of cells that are most highly correlated or anti-correlated with past performance. P-values for the positive and negative tail of the distribution are shown at the top (2 sample variance test). **e-f,** distributions of correlation to learning curve for heat on (**e**) and heat off (**f**) activity. Original dataset is shown in black, scrambled dataset (by trial order) is shown in red. Scrambling by trial order reduces number of cells with strong correlation or anti-correlation to the learning curve for heat on activity, but only reduces the number negatively correlated heat off cells. Consequently, heat off signal with positive correlation to the learning curve and past performance were not analyzed. P-values for the positive and negative tail of the distribution are shown at the top (2 sample variance test). **g-h**, distribution of error/correct trial specific heat on activity (**g**) and heat off activity (**h**). Each neuron is scored between -1 and 1. A score of -1 means that all activity occurred within error trials, and a score of 1 means that all activity occurred within correct trials. Scrambling by trial order (red) decreases the number of neuron with error trial specific activity. P-values for the positive and negative tail of the distribution are shown at the top (2 sample variance test).

To determine whether the observed correlations to past performance and differential error/correct trial activity could be explained by random fluctuations in neuronal activity across a large population of neurons, we used two scrambling approaches to determine whether our data contain statistically significant signals. In the first scrambling approach, we circularly permuted our signal relative to sensory and motor event triggers. This scrambling approach decreases the number of cells with stimulus specific and no ITI activity by 5.3 fold (1-ANOVA, $p < 0.0001$) (**Fig. 8a**).

To more finely test whether the correlation to learning and past performance could be the result of random sensory and motor event-related activity, we performed a more conservative scramble, in which we maintain the relationship of each neuron to its sensory and motor triggers, but randomize the trial order to abolish the relationship between any learning-related signals and the learning process. This scramble does not change the distribution of stimulus specific activity (**Fig. 8b**), but significantly reduces correlation to past performance and differential activity between correct/error trials (**Fig. 8c-h**). The scrambled dataset shows an increased number of cells with no correlation to learning, and truncates the tails of the correlation coefficient distributions representing strong positive and negative correlation to learning. Correspondingly, the scrambled dataset shows a greater number of cells with indistinguishable error/correct trial activity and fewer cells with error or correct trial-specific activity.

We then designed a set of mutually exclusive classifiers to identify cells that encode sensory, motor, and learning-related activity (see methods). We define learning related activity as activity that is modulated by past performance. Past performance is defined as the fraction of previous trials that are correct (**Fig. 9**). This metric is closely related to

correlations to the fitted learning curve, but makes fewer assumptions about the rate of learning and stability of the learned response. For each type, we test our classifier against both scrambling methods to determine the statistical significance of the identified classes. All sensory, motor and learning-related classes should be significantly reduced by the first scrambling method, while only learning-related activity should be disrupted by the second scrambling method. We will limit our discussions to functional types that have been shown to be statistically significant based on the relevant scrambling methods (1-way ANOVA, $p < 0.0001$), and have a false discovery rate of under 0.3.

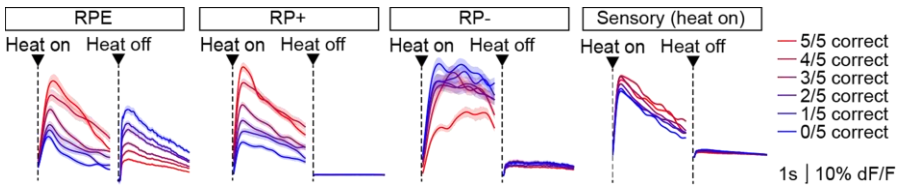

Figure 9 | Neuronal classes modulated by past performance. Heat onset and offset activity in four classes of neurons. RPE, RP+, and RP- classes show modulated activity as a function of the fraction of trials that were correct out of the last five trials, whereas the sensory heat on neurons are relatively insensitive to past performance. Error bars represent s.e.m.

We find three neuronal classes that are modulated by past performance in both training blocks (**Fig. 9**). The activity of these three groups of neurons suggests that they may encode relief prediction error (RPE), heat onset with relief prediction (RP+), and heat onset without relief prediction (RP-). RP+ and RP- neurons are positively (0.36 ± 0.12, s.d.) and negatively (-0.26 ± 0.11, s.d) modulated by past performance, respectively. Activity of the RPE neuron contains both heat on and heat off components. The heat on

component of RPE neurons are positively (0.29 $\pm$ 0.17, s.d.) modulated by past performance, while the heat off component is negatively (-0.18 $\pm$ 0.14, s.d.) modulated.

Over the course of the first training block, RPE neurons are initially activated by heat offset, but shift its activity toward heat onset as learning progresses (**Fig 10a-b**). By fitting the behavior to a sigmoidal learning curve, we can define a transition point that divides the block into pre-learning trials and post-learning trials. By the end of the first training block, the post-learning heat on response of an RPE neuron is significantly higher than its pre-learning heat on response, while the response to heat offset has undergone the opposite pattern. During the first trial of the reversal block, we observe that the initial response to heat onset remains high, since the expectations of the animal have not yet updated even though the reward contingency has changed. By the second trial of the reversal block, we begin to observe a decrease in response to heat onset. This decrease in activity is reversed as the animal begins to learn the new reward contingency. In contrast to heat onset activity, we observe strong heat offset activity in the first reversal trial, as the heat terminates unexpectedly following a previously incorrect turn. As the animal begins to learn the new reward contingency, activity again shifts from heat offset to heat onset. This shift in activity from relief (heat offset) to a sensory event that consistently precedes relief (i.e. heat onset) is a signature of prediction error signals.

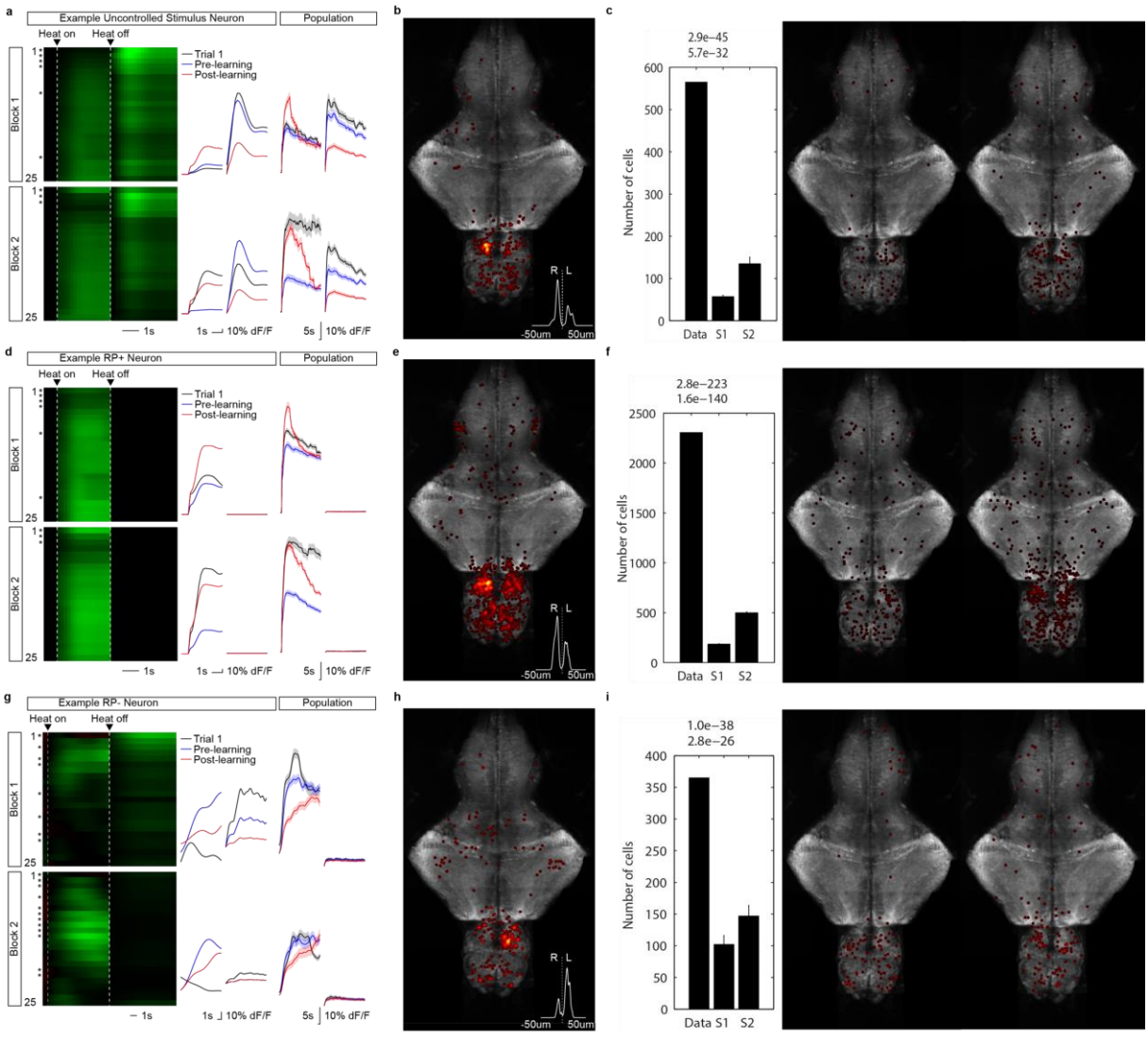

Figure 10 | Prediction-related signals in the habenula and pallium. a,d,g, Single-cell analysis reveals three classes learning-related activity that may encode relief prediction error (**a**, RPE), heat onset with relief prediction (**d**, RP+), and heat onset without relief prediction (**g**, RP-). Activity rasters (left) show activity of a representative neuron in each class following heat onset and offset (arrows and dotted white lines) across trials in block 1 and block 2. Example traces (middle) show the activity of the same neuron in the first trial of each block (black lines), averaged in trials before the learning transition (blue lines), and averaged in trials after the learning transition (red lines). Population summary traces (right) show the same signals, averaged across all neurons identified by our classifiers (see Methods). Error bars represent s.e.m. **b,e,h,** Anatomical locations of the prediction-related classes defined in **a,d,g**. Insets summarize the distribution of each class across the left/right axis of the animal.

Figure 10 | Prediction-related signals in the habenula and pallium (continued). c,f,i, The effect of two scrambling methods on classification of RPE (**c**), RP+ (**f**), RP- (**i**). For each type, the number of cells found from the original data (data), scrambled dataset by circular permutation (S1), and scrambled dataset by randomizing trial order (S2) are shown from left to right. Standard deviation is shown for each scrambling method. P-values (1-way ANOVA) for the reduction of cell number by S1 and S2 are shown at the top of the bar plot, respectively. Distribution of found cells from S1 and S2 are shown beside the bar plot from left to right, respectively.

In the context of reward learning, the temporal difference model proposes a link between neurons that encode reward prediction error and neurons that neurons that encode reward prediction following a neutral reward-predicting cue. It is unknown whether similar processes govern relief prediction following the presentation of an aversive cue. Our results suggest relief prediction error and RP+ neurons are in fact linked in an analogous fashion (**Fig. 10d-e**). RP+ neurons are only activated by heat onset and are positively modulated by past experience (**Fig. 9**). Over the course of the first training block, these cells increase their activity as learning progresses. During the first reversal trial, the activity in response to heat onset remain highs, since the animal continues to expect relief. As error trials accumulate during the pre-learning phase of the reversal block, RP+ activity is suppressed. As the animal learns the new reward contingency, RP+ activity is reinstated. The changes in activity of the RP+ neuron mirror the corresponding changes in response latency during learning. As past performance improves, RP+ neurons become increasingly active, while motor response latency becomes diminished.

In contrast to both RP+ and RPE neurons, we observe a third class of neurons that are activated by heat onset, but whose activity is negatively modulated by past performance

(RP-). These neurons increase both their peak activity and slope of response as a function of the fraction of previous trials that were error trials (**Fig. 10g-h, Fig. 9**). Initially, these neurons respond to heat onset, but as the learning progresses and the number of error trials decreases, these neurons suppress their activity in response to heat onset. During the first reversal trial, the activity is initially low, and then increases rapidly after the first error movement. As error trials accumulate during the pre-learning phase of the reversal block, RP- neurons begin to respond more strongly and rapidly to heat onset. As the animal learns the new reward contingency, the activity of these neurons again becomes suppressed. The RP- neurons appear to encode noxious heat in the absence of relief prediction. These neurons may be related to cells that show increased activation during uncontrollable stress or inescapable shock paradigms. To our knowledge, it has not been previously shown for these paradigms whether uncontrollable stress related signals are active during normal learning or whether their activity is specific to a relief prediction context.

To further test whether these three classes of cells reflect prediction-related activity, we surveyed their activity in an open loop block. During open loop, the heat terminates after a random interval (5-30s) for each trial, rendering relief unpredictable. Behaviorally, the open loop condition leads to loss of turn direction bias and increased response latency **Fig. 5)**. Consistent with this change in response latency, we find that RP+ neuron decrease in activity over successive trials while RP- neurons increase in activity over successive trials. RPE neurons show increased heat off activity and decreased heat on activity (**Fig. 11**).

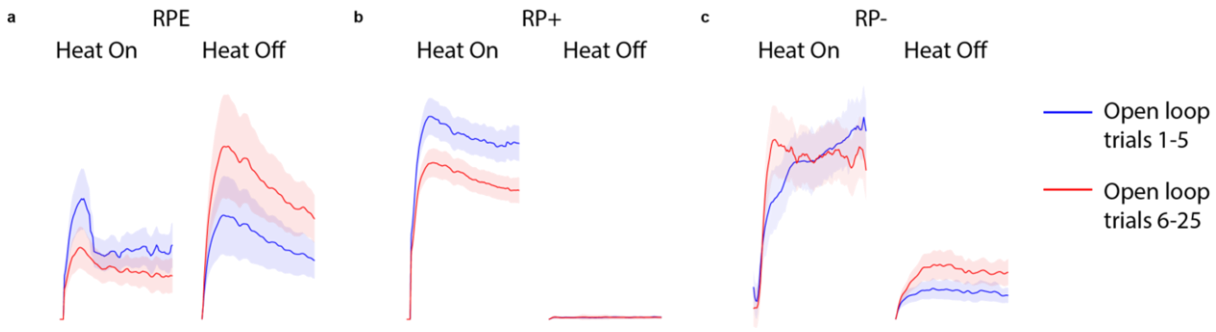

Figure 11 | Activity of RPE (a), RP+ (b) and RP- (c) neurons in open loop. Mean heat on and off activity from the first five trials of open loop are shown in blue. Mean heat on and off activity from the trial 6-25 of open loop are shown in red

Lastly, we examine the activity pattern of these three cell types in response to the first error movement of the reversal trial (trial 1, block 2) (**Fig. 12**). This trial effectively serves as an omission trial, in which expected relief is unexpectedly withheld. In a subset of RPE neurons (26%), we observe a clear inhibition of response after the first movement (see discussion). For RP+ neurons, the peak response occurs around the time of the first movement and then decays, while for RP- neurons, the peak response occurs after the first movement.

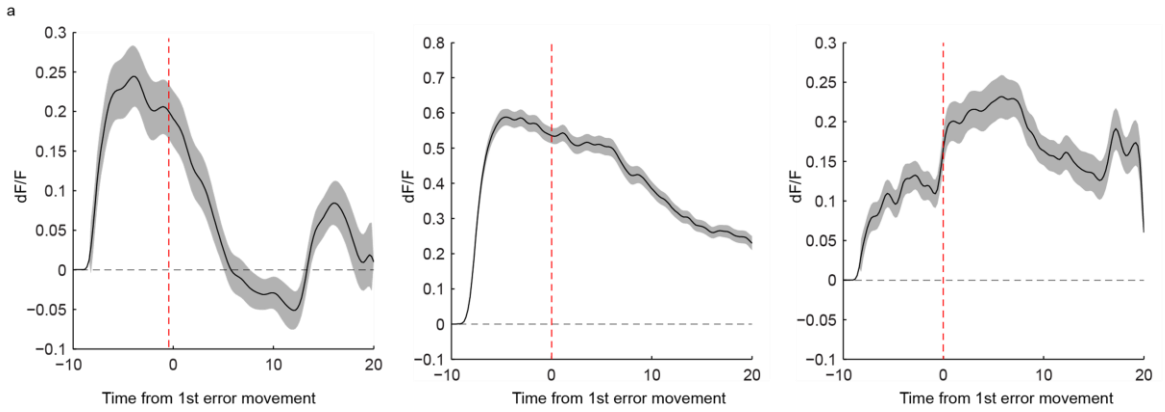

Figure 12 | Reversal trial activity triggered around the first error turn. a, Average of a subset of RPE neurons (26%) that show inhibition after the first error movement (red dotted line) in reversal trials. **b,** Average of all RP+ neuron activity around the first error movement. **c,** Average of RP- activity around the first error movement. (A reversal trial is defined as the first trial of the reversal block.)

Anatomical distribution of prediction-related signals

RPE, RP+, and RP- signals are all highly enriched in the habenula, and show a striking lateralization (**Fig 10b,e,h**). RPE and RP+ occupy the right habenula, with a 2-fold and 1.4-fold enrichment over the left habenula respectively. In contrast, the RP- signal is tightly localized to the left habenula (3.8-fold enrichment over the right habenula). Datasets generated by either scrambling methods do not show a similar pattern of localization (**Fig. 10c,f,i**). The segregation of functional types suggests a lateralization of habenula function in learning that has not been previous described. Though the habenula has been shown to be anatomically and genetically asymmetric across the left-right axis [48, 72, 130-133], electrophysiological studies of habenula function in mammals have not routinely compared activity across the left and right axis[47-67].

In contrast to the relief prediction signals described above, heat on and heat off signals that have low (<0.1) absolute correlation to past performance show a much wider

anatomical distribution across the forebrain, midbrain, and hindbrain (**Fig. 13**). These signals appear to simply relay the thermosensory stimulus, and show no significant lateralization across each brain region.

Of all the functional classes we examined, we do find another class that shows significant lateralization across the left and right habenula – a reversal signal[14-16]. Cells of this type are most active during the first several trials following block reversal, and then as learning progresses, their activity is suppressed (**Fig. 14**). These signals may represent a variant of the RP- signal that is activated specifically when a previously predictive cue becomes unpredictable. Consistent with this view, these signals also show greater enrichment in the left habenula.

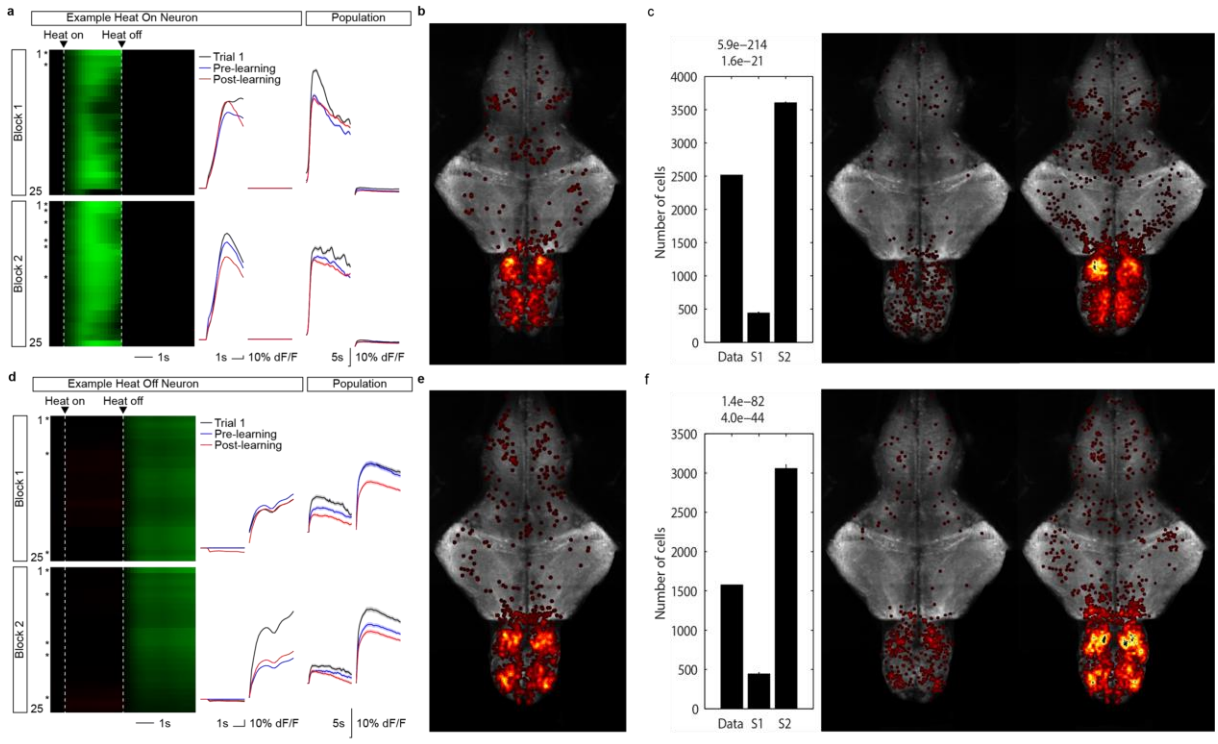

Figure 13 | Heat onset-related and heat offset-related sensory classes. a,d, Single-cell analysis reveals heat onset-related (**a**) and heat offset related (**d**) classes that are relatively insensitive to learning progression. Activity rasters (left) show activity of a representative neuron in each class following heat onset and offset (arrows and dotted white lines) across trials in block 1 and block 2. Example traces (middle) show the activity of the same neuron in the first trial of each block (black lines), averaged in trials before the learning transition (blue lines), and averaged in trials after the learning transition (red lines). Population summary traces (right) show the same signals, averaged across all neurons identified by our classifiers (see Methods). **b,e,** Anatomical locations of the classes defined in **a,d**. Insets summarize the distribution of each class across the left/right axis of the animal. Error bars represent s.e.m. **c,f,** The effect of two scrambling methods on classification of heat on neurons (**c**) and heat off neurons (**f**). For each type, the number of cells found from the original data (data), scrambled dataset by circular permutation (S1), and scrambled dataset by randomizing trial order (S2) are shown from left to right. Standard deviation is shown for each scrambling method (3X scramble). P-values (1-way ANOVA) for the reduction of cell number by S1 and S2 are shown at the top of the bar plot, respectively. Distribution of found cells from S1 and S2 are shown beside the bar plot from left to right, respectively.

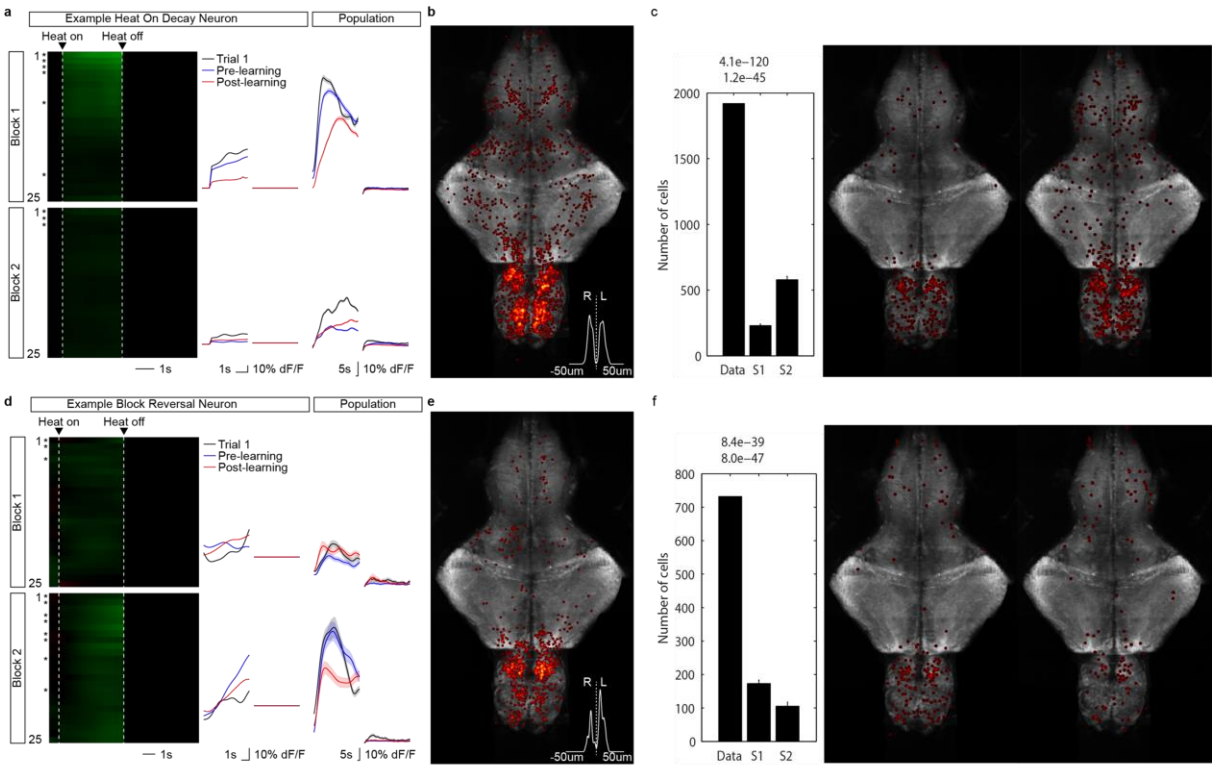

Figure 14 | Heat on decay and reversal-related classes. a,d, Single-cell analysis reveals heat onset-related neurons that decay across trials throughout the paradigm (**a**) and heat onset-related neurons that are differentially more active during reversal (**d**). Activity rasters (left) show activity of a representative neuron in each class following heat onset and offset (arrows and dotted white lines) across trials in block 1 and block 2. Example traces (middle) show the activity of the same neuron in the first trial of each block (black lines), averaged in trials before the learning transition (blue lines), and averaged in trials after the learning transition (red lines). Population summary traces (right) show the same signals, averaged across all neurons identified by our classifiers (see Methods). **b,e,** Anatomical locations of the classes defined in **a,d**. Insets summarize the distribution of each class across the left/right axis of the animal. Error bars represent s.e.m. **c,f,** The effect of two scrambling methods on classification of heat on decay neurons (**c**) and heat off decay neurons (**f**). For each type, the number of cells found from the original data (data), scrambled dataset by circular permutation (S1), and scrambled dataset by randomizing trial order (S2) are shown from left to right. Standard deviation is shown for each scrambling method (3X scramble). P-values (1-way ANOVA) for the reduction of cell number by S1 and S2 are shown at the top of the bar plot, respectively. Distribution of found cells from S1 and S2 are shown beside the bar plot from left to right, respectively.

Single fish analysis of relief prediction-related signals

If the RPE, RP+, and RP- signals participate in a circuit that exhibit opponency across the left and right habenula, we would expect to find all three signals co-existing in a single animal. Alternatively, it could be that individual animals only possess a subset of the three signals, and the functional asymmetry that we observe in the population data are not representative of the neural circuit in an individual animal. To address this potential caveat, we analyzed activity across a single habenula that contained all three prediction related signals (**Fig. 15**).

Since these three types of signal show distinct patterns of activity across trials, we hypothesized that we would be able to independently recover these three types based on an unbiased cross correlation analysis across all cells (**Fig 15a-b**). We find that out of 359 of neurons, 72 neurons shown a >0.5 correlation in activity with at least one other cell. PCA analysis of this subset of neurons reveals three distinct clusters, with 33, 23, 11 cells in each cluster (**Fig. 15c**). These clusters can also be identified by hierarchical clustering. (6 neurons did not cluster with the three largest clusters by hierarchical clustering, and were not included the subsequent analysis.) When we averaged the activity of each cluster, we find that the activity of the three clusters correspond to RP+, RP-, and RPE (**Fig. 15f-h**).

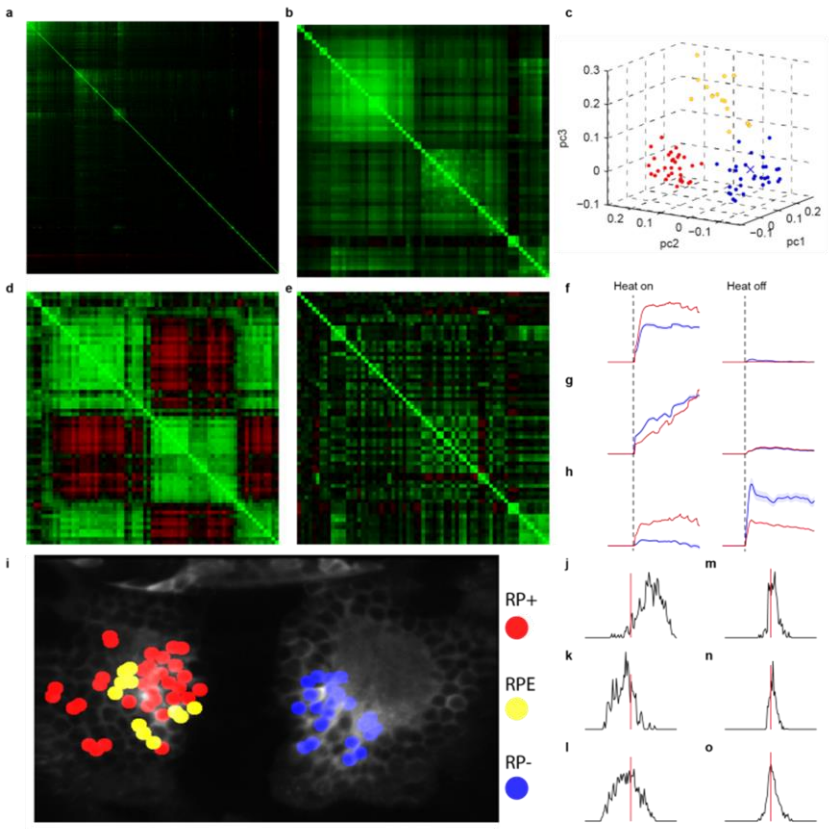

Figure 15 | Single-fish analysis of relief prediction-related signals. a, Activity correlation matrix for all neurons in a dataset of 359 neurons imaged simultaneously in the habenula of a single fish. Rows and columns are sorted by hierarchical clustering to reveal 72 neurons with strong correlation to at least one other neuron. **b,** Zoom of the three clusters of neurons with strong correlation to other neurons in the dataset. **c,** PCA analysis of the correlation vectors for these 72 neurons confirms that the three clusters identified by hierarchical clustering are distinct. **d-e,** Activity correlation matrix of these 72 neurons using only their heat on-related (**d**) or heat off-related (**e**) activity shows that all three clusters are correlated or anticorrelated during the heat on period. The smallest cluster is positively correlated during the heat off period. **f-h,** Average activity traces of each cluster in response to heat onset and heat offset in trials before learning (blue) and after learning (red). The three clusters correspond to RP+ (**f**), RP- (**g**), and RPE (**h**) neurons. **i,** Anatomical locations of the RP+, RP-, and RPE clusters. RP- neurons are located in the left habenula, while RPE- and RPE neurons are located in the right habenula. **j-o,** Analysis of correlation between RP+ and RPE neurons (top), between RP+ and RP- neurons (middle), and between RP- and RPE neurons (bottom). Plots show the distribution of correlation coefficients across all such pairs, using activity recorded during the learning paradigm (left) or spontaneous activity recorded during 2 min before the paradigm (right). Red vertical lines mark correlation coefficient of 0.

We then separated each activity trace by its heat on and heat off component, and repeated the cross correlation analysis for the three clusters (**Fig. 15d-e**). For the heat on component, we find strong positive correlation (0.43 $\pm$ 0.012, s.e.m.) between RP+ and RPE neurons, and negative correlation (-0.15 $\pm$ 0.012, -0.08 $\pm$ 0.009, s.e.m.) between RP+/RPE and RP- neurons (**Fig. 15j-l**). For the heat off component, we find the highest correlation (0.52 $\pm$ 0.027, s.e.m.) within the RPE cluster. These results are consistent with our view that RP+ and RP- represent opposing prediction-related states. We also find a small but consistent positive correlation in spontaneous activity across all three clusters (**Fig. 15m-o**), which may suggest some shared input (i.e. sensory heat on) to all three clusters.

Finally, we examined the anatomical localization of all three clusters, and find a clear segregation across the left and right habenula (**Fig. 15i**). All RP+/RPE neurons are localized to the right medial habenula, while all RP- neurons are localized to the left medial habenula. This segregation recapitulates the pattern of lateralization observed across all RP+, RPE, and RP- neurons.

Implementation of the learned response: action selection

Both the sensory and prediction-related cells described above are not selective for reward contingency. In order to bias motor output toward the block-specific correct response, there must be neurons that discriminate between the reward contingencies of the two blocks. These neurons could be involved in representing or driving the correct action in a given block, and so we refer to them as action-related neurons[23-35]. We will show that these action-related neurons can be distinguished from motor-related

neurons that are tightly associated with the execution of individual turn events independent of context or reward contingency. To identify action-related and motor-related neurons, we use a combination of peak activity and ROC analysis for each neuron (see methods).

Action-related neurons are distinct from motor-related neurons in several ways (**Fig. 16**). 1) Motor-related neurons are active during both spontaneous turns and turns evoked during the heat on period. Left motor-related neurons are active before both spontaneous and heat- driven left turns, and right motor-related neurons are active before both spontaneous and heat-driven right turns (**Fig. 16c-d**). In contrast, left action-related neurons are most strongly activated before heat-driven left turns (**Fig. 16a**), and right action-related neurons are most strongly activated before heat- driven right turns (**Fig 16b**). Action-related neurons show little activation before or after spontaneous movements. 2). Motor-related neurons are active immediately before the actual turn (< 1s), whereas action-related neurons can show significant activity >10s before the actual turn. An action-related neuron can predict the direction of the next heat driven turn many seconds before motor-related neurons. 3) The anatomical distribution of action-related and motor-related neurons are clearly distinct. Motor-related neurons are found throughout the forebrain, midbrain and hindbrain (**Fig. 16g**). Right motor-related neurons are enriched in the right medial hindbrain and midbrain, while left motor-related neurons are enriched along the left medial hindbrain and midbrain. In contrast, action-related neurons are found predominantly in the forebrain, in both the habenula and pallium (**Fig. 16e-f**).

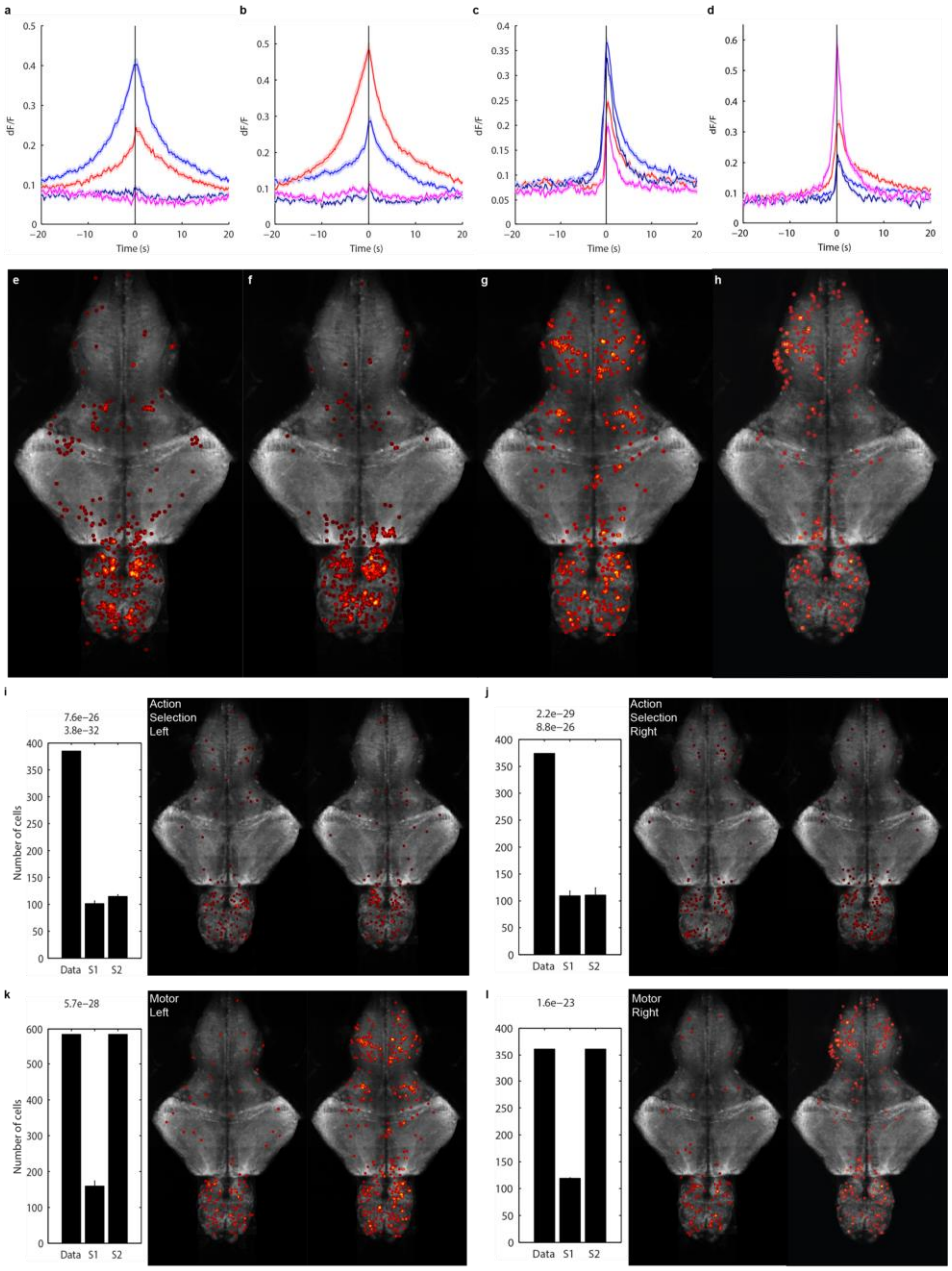

Figure 16 | Action selection neurons and motor-related neurons. a-d, Classifiers based on peak activity and ROC analysis identify classes corresponding to left action selection (**a**), right action selection (**b**), left turns (**c**), and right turns (**d**). Plots show average activity after aligning traces to the time of turn events, computed separately for heat-evoked left turns (red lines), heat-evoked right turns (blue lines), spontaneous left turns (pink lines), or spontaneous right turns (purple lines). **e-f,** Anatomical locations of left action selection neurons (**e**), right action selection neurons (**f**), left turn neurons (**g**), and right turn neurons (**h**). Error bars represent s.e.m.

Figure 16 | Action selection neurons and motor-related neurons (continued). i-l, The effect of two scrambling methods on classification of left action selection (**i**), right action selection (**j**), left turn motor-related (**k**), and right turn motor-related (**l**) neurons. For each type, the number of cells found from the original data (data), scrambled dataset by circular permutation (S1), and scrambled dataset by randomizing trial order (S2) are shown from left to right. Standard deviation is shown for each scrambling method (3X scramble). P-values (1-way ANOVA) for the reduction of cell number by S1 and S2 are shown at the top of the bar plot, respectively. Distribution of found cells from S1 and S2 are shown beside the bar plot from left to right, respectively.

Consistent with the hypothesis that action-related neurons bias the animal toward the learned turn direction, we find that left action-related neurons are differentially suppressed in right turn rewarded blocks, and differentially active in left turn rewarded blocks, while the opposite pattern holds for right action-related neurons. On average, left action-related neurons show positive correlation with learning (0.06 ± 0.01, s.e.m.) in left rewarded blocks, and a negative correlation with learning (-0.13 ± 0.1, s.e.m.) in right rewarded blocks. In contrast, right action-related neurons show a negative correlation with learning (-0.12 ± 0.02, s.e.m.) in left rewarded blocks, and a positive correlation with learning (0.03 ± 0.02, s.e.m.) in right reward blocks.

For the action-related neurons described above, we found that if the reward contingency of the first training block is against the preferred direction of an action-related neuron, then the neuron's activity is suppressed during the first block and reactivated during the second block. We then asked whether all cells that are suppressed in the first training block are reactivated in the second training block. We find instead a large subset of heat activated neurons that decay in activity over the first training block and are not reactivated in the subsequent block (**Fig. 14a**). This class of neurons is particularly enriched in the medial pallium.

Non learners lack RP- and action selection signals

Are the neurons that encode prediction and action selection in fact necessary for learning? We use the natural variability in behavior to address this question. At 6-7dpf, 28% of animals fail to learn at least one of the two reward contingencies. We classify these animals as non-learners. Network activity in non-learners lacks the stereotyped pattern of rapid transition and decay to steady state observed in learners (**Fig 17a**). We then ask whether specific functional types are significantly disrupted in non-learners. We find that the most severely disrupted classes are RP-, action selection, and reversal signals (**Fig. 17b**). These three classes were reduced by 75% ($p < 6.7e-08$), 77% ($p < 1.1e-26$), and 43% ($p < 4.1e-05$), respectively. Both RP- and reversal signal are activated by heat onset when relief is uncertain. One possible function for this type of signal is to inform the animal that the current behavior is no longer adaptive, and facilitate switching to new behavior programs (e.g. by recruiting previously inactive action selection neurons). In the absence of these signals, the animal would be unable to adapt to new reward contingencies. The absence of action selection neurons due to either lack of recruitment or reinforcement by the RPE signal could then account for the lack of correct turn bias.

The total number of sensory-related, RP+, RPE neurons show only modest decreases in non-learners by 22% ($p < 1.3e-05$), 37% ($p < 1.1e-05$), and 29% ($p < 0.0045$), respectively. The relative distributions of these signals are also preserved in the non-learners. Both RP+ and RPE signals show enrichment in the right habenula in both learners and non-learners. In contrast, any remaining RP- and action selection cells in non-learners show no apparent asymmetric distribution in the habenula. We also

observe an increase in the number of heat on-related decay neurons, which may also reflect the lack of reactivation during the reversal block in non-learners.

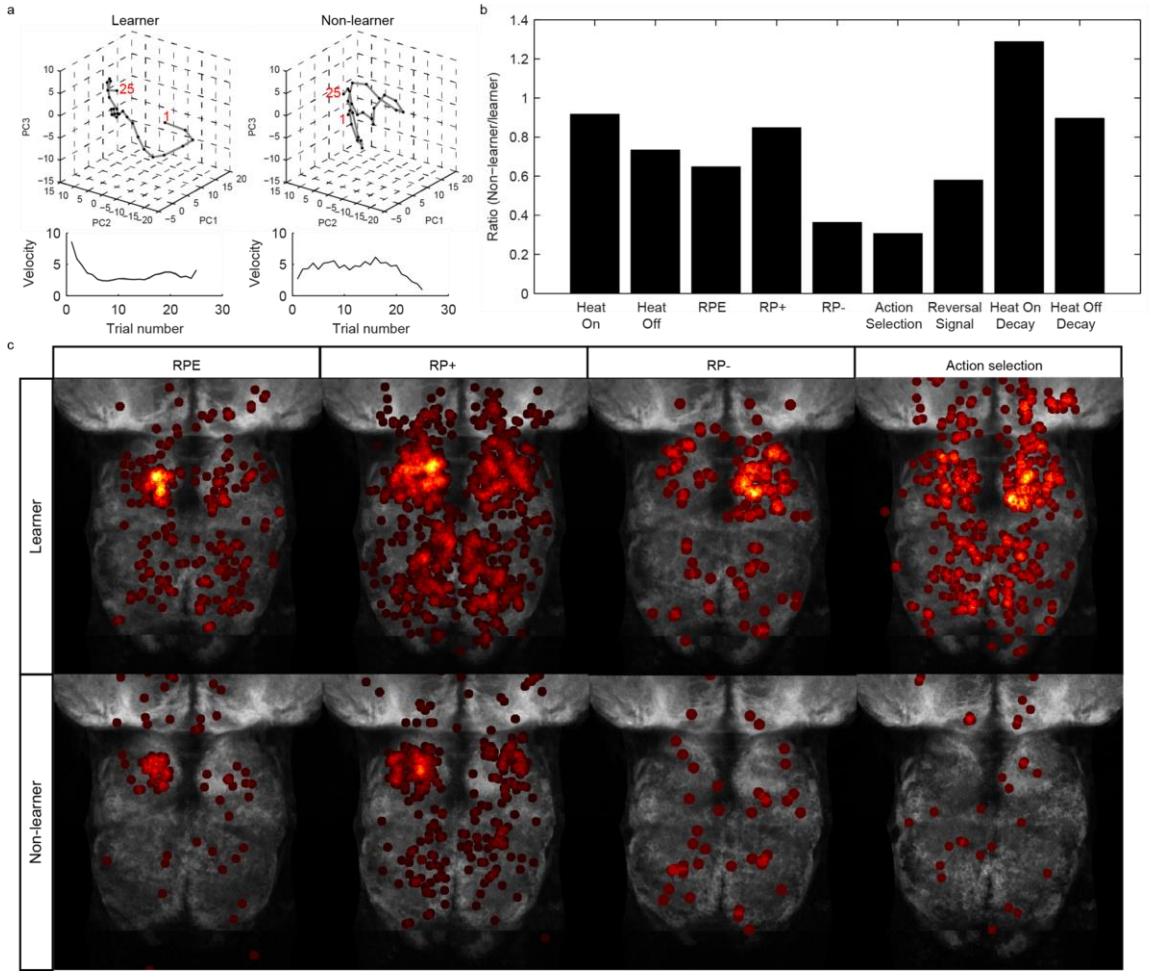

Figure 17 | Non-learners lack RP- and action selection neurons. a, Sample PCA trajectory and average PCA trajectory velocity of all Habenula neurons from learners and non-learners. **b**, Ratio of the number of forebrain neurons found in non-learner over learners for each functional class. Ratios were normalized by the total number of active neurons found in learners (10383) and non-learners (5805) at 6-7 dpf. Animals with switching deficit were excluded from the analysis. **c**, Distribution of RPE, RP+, RP-, and action selection neurons in the forebrain for learners and non-learners.

Global analysis of network activity reveals learning-related changes

We first take a global approach to determine whether learning generates detectable changes in network activity. For each brain region, we used principal component analysis to construct a trajectory representing the evolution of population activity state across trials (**Fig. 18a-c).** We analyzed the velocity of a given circuit trajectory as an approximation of the rate of activity changes across a given brain region (**Fig. 18d-f**). We find that the circuit trajectories of regions such as the habenula, pallium, subpallium, and hindbrain show the largest changes in neuronal activity across successive trials within each training block. In contrast, midbrain structures such as the optic tectum do not show consistent rates of activity change across both blocks.

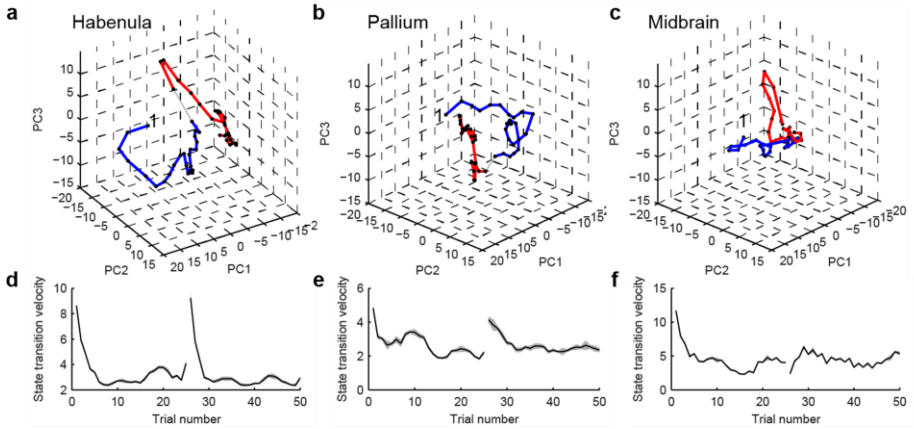

Figure 18 | PCA trajectory and velocity c-e, Representative trajectories showing the evolution of network activity state across trials in block 1 (blue) and block 2 (red), using randomly sampled subpopulations of 500 neurons in the habenula (**a**), pallium (**b**), and midbrain (**c**). Network activity state was projected onto the first three components (PC) obtained by PCA. **d-f,** Average rate of change of population state across trials in each block, using randomly sampled subpopulations of 500 neurons in the same brain regions as in **a-c**. The habenula and pallium undergo rapid state changes in the early trials of both blocks, whereas the midbrain does not show consistent changes across both blocks. Error bars represent s.e.m. based on 100 replicates of randomly sampling subpopulations of 500 neurons from each brain region. The entire dataset consists of 24,546 task-related neurons.

Of all brain regions, the habenula shows the largest rate of activity change across both training blocks. The rate of change is highest between the first and second trials of each block, and exponentially decays toward a steady state within the first five trials. This indicates that the largest shift in neural activity occurs during the early trials of each training block. The rate of activity change is highly correlated with the latency to correct response (**Fig. 1d**), which also shows the largest decrease between the first two trials of training and exponentially decays to a steady state after 6 trials.

To further test whether the rate of activity change is related to learning, we repeated the PCA analysis for single animals. We find that an animal that required more error trials to learn also required more trials to enter the steady state phase of the circuit trajectory (**Fig. 19**). Overall, there is a positive correlation between the learning curve of an animal and the rate of neuronal activity change (0.11). This suggests that the steady state phase of the circuit trajectory represents a post-learning network state, while the initial transitional phase reflects the learning process.

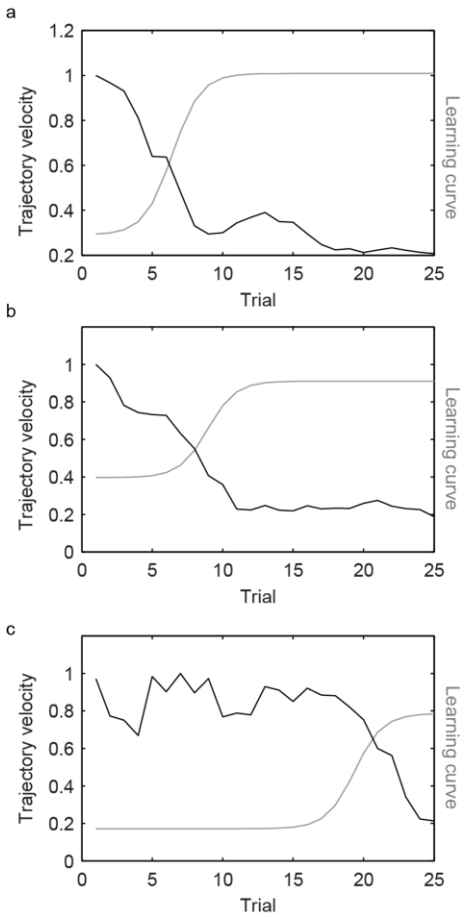

Figure 19 | Examples showing the correspondence between learning curve (gray) and trajectory velocity (black). For each fish, the learning curve was obtained by fitting a sigmoid function to sequence of correct and incorrect trials.

To determine whether the changes in activity state can be observed for either reward contingency, we repeated the PCA analysis, except that we grouped trials across fish according to the reward contingency of the trial rather than the order in which the blocks were collected (**Fig. 20**). We find a similar rate of activity change for both reward contingencies. We also note that the two circuit trajectories occupy separate state space, which may reflect the presence of reward contingency-specific activity (see next section).

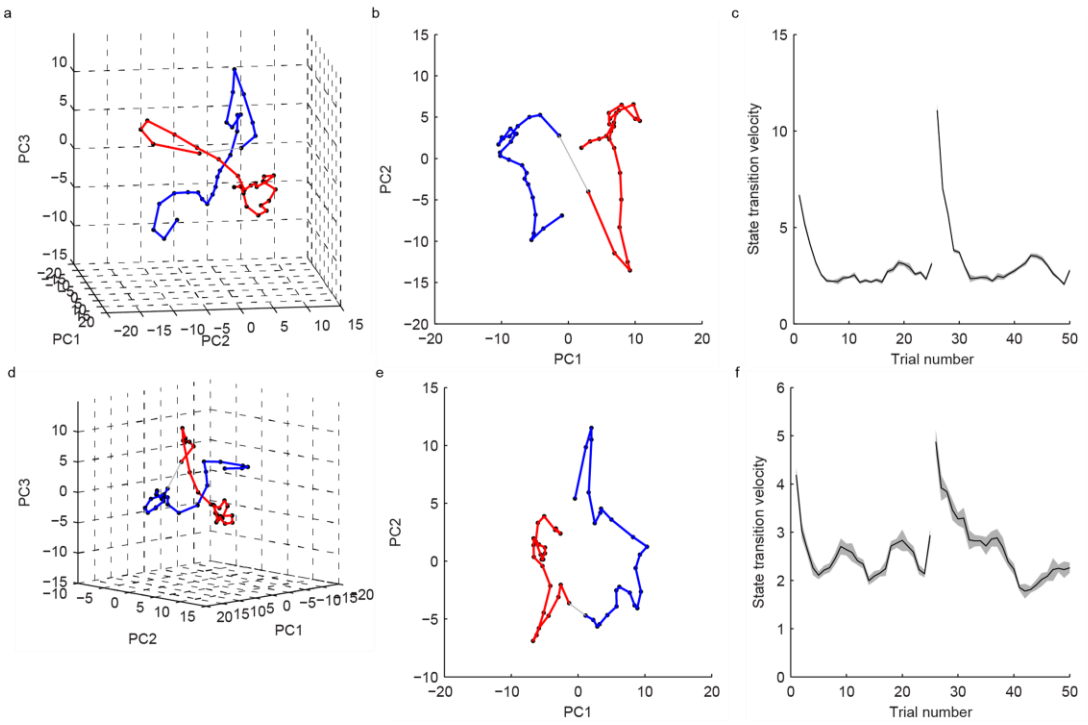

Figure 20 | PCA trajectory and trajectory velocity for the habenula (a-c) and pallium (d-e) by reward contingency. Trajectories from left rewarded blocks are shown in blue, and trajectories from right reward blocks are shown in red.

To control for the potential caveat that the changes in network activity simply reflect changes in sensory experience and motor output, we excluded all motor-related neurons and restricted our analysis to only a small time window around the start of heat onset (**Fig. 21**). We used 10 s before heat onset and 5 s after heat onset, excluding activity that occurred after the first movement. This restricted time window precludes contribution from activity that occurred only in response to error movements and activity that is sensitive to the duration of heat exposure. We wanted to avoid these types of activity because although they would increase the variance between correct and error trials, these signals might simply reflect sensory rather than learning-related changes

across trials. Performing PCA on this restricted dataset, we still find that the habenula shows rapid shifts in activity state in the early trials of each training block. In contrast, the hindbrain now no longer shows consistent shifts in network activity for both blocks. Interestingly, the pallium shows a much larger shift in activity during the first training block than the reversal training block. This difference between the habenula and pallium is reflective of the relative proportion of block 1 and block 2-specific signals found in these two regions. We found that the medial pallium contains a large number of heat activated neurons that decay in activity over the first training block and are not reactivated in the subsequent block. This class may account for the large changes in network state observed in block 1 of PCA trajectory analysis for this region.

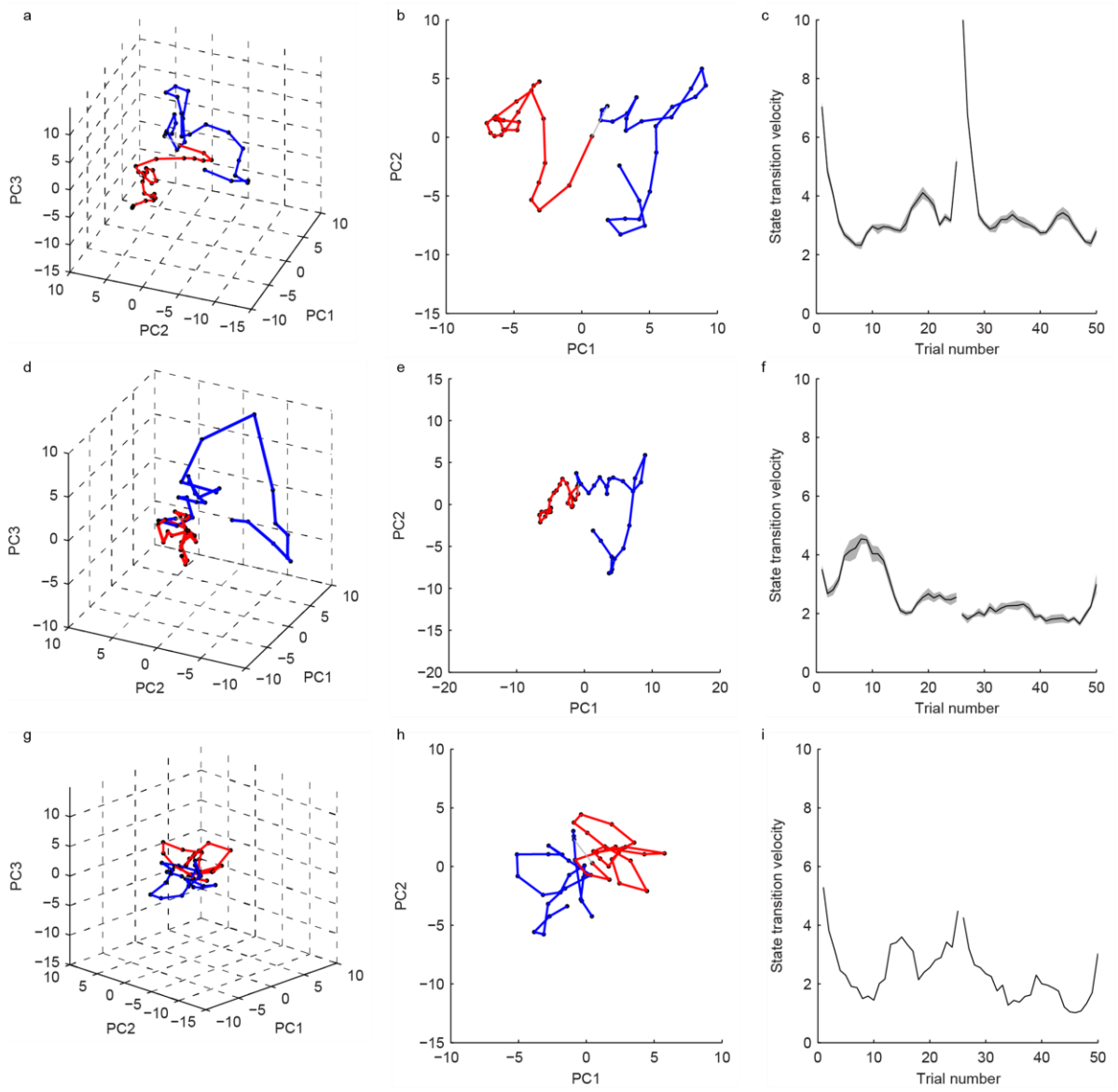

Figure 21 | PCA trajectory and trajectory velocity for habenula (a-c), pallium (d-f), and hindbrain (g-i) with restricted activity window. For all three brain regions, motor neurons were not included in the analysis and PCA was performed on a restricted window of activity for each trial (-10s before heat onset and 5s after heat onset). All activity occurring after the first movement was excluded. Block1 is shown is blue and block2 is shown in red.

Conclusion and Discussion

We have developed a novel paradigm to assay operant avoidance learning in larval zebrafish while exploiting its unique advantages for functional imaging at single cell resolution throughout the brain. Our analysis of the brain activity of more than 200 animals during learning reveals a diverse array of functional types, ranging from simple sensory and motor signals to action-related signals and a variety of prediction-related signals that may encode relief prediction error (RPE), as well as heat with or without predicted relief (RP+ and RP-, respectively). In addition, our analysis of the anatomical distributions of these functional classes has revealed that most learning-related activity in our operant task is localized to the habenula and forebrain, and that some classes show intriguing functional lateralization.

Examination of individual neurons with correlations to past performance reveals a highly lateralized relief prediction circuit in the habenula. Neurons that encode relief prediction and relief prediction error appear to be localized to the right medial habenula, while neurons that encode the absence of relief prediction are localized to the left medial habenula. Although the habenula has been shown to be high asymmetric both anatomically and in terms of gene expression[48], learning-related asymmetry has not been described. Our results suggest that the left and right habenula represent opponent processes in relief prediction, and it remains to be seen whether this functional asymmetry is conserved in other vertebrates.

In mammals, dopaminergic RPE neurons mediate operating learning by interacting with action control neurons in the striatum[23-35]. We find a number of neurons that may

mediate action selection in our behavior paradigms. These neurons are selectively active before heat-evoked left or right turns. As a population, action selection cells can predict the direction of the next heat-evoked movement more than 10 s in advance of the actual movement. In non-learners that fail to learn the correct response for a given reward contingency, significantly fewer action selection neurons are observed for the rewarded turn direction. This disruption is also observed for RP- but not RP+ and RPE cells. In the future, ablation experiments will be necessary to determine whether RP- neurons are necessary to recruit action selection neurons, or whether the non-learners lack direct synaptic connection between RPE and action selection neurons.

The absence of left habenula signals (RP- and reversal signal), may be also point to developmental delays as a contributing factor to the inability to learn in a subset of animals. Left habenula develops more slowly than the right habenula by anatomy, gene expression, and functional activity (see next chapter). Though the size of the left habenula is not significant different between learners and non-learners at 6-7 dpf, more subtle delays in synaptic connectivity may give rise to the observed behavioral and functional phenotypes. Viral tracing and/or EM reconstruction will help to address this question in the future.

We note that only a subset of our RPE neurons show inhibition after the first error movement in reversal trials (block 2, trial 1). There are several possible explanations. Inhibition is difficult to detect in calcium imaging unless the baseline firing rate is relatively high. Otherwise, slight decreases in calcium activity are not likely to rise above random fluctuation in baseline fluorescence. Furthermore, the long decay constant of the calcium signal can also serve to mask small inhibitory responses that follow an initial

increase in activity. These issues may be addressed with faster and brighter calcium indicators (e.g. GCaMP6f), voltage indicators, or electrophysiological recordings. However, all three methods are still under active development for the larval zebrafish system.

Another fascinating complexity is that a subset of RPE neurons appears to respond before each error movement in the reversal trial. In an operant learning paradigm, the cue for reward/relief can be either a sensory stimulus or activity of action selection neurons[137]. In other words, it is possible for an animal to predict reward based on an upcoming behavioral choice. This type of RPE neuron would fire before each error movement in the reversal trial, though the peak response would decrease as the animal updates the prediction value of the action. If this were the case, then inhibition would again be difficult to observe, since our paradigm does not control how many error turns an animal makes or the duration between error trials. In future experiments, we could introduce a delay between the movement and the eventual outcome to more clearly assay the activity of RPE neurons in omission or reversal trials.

Methods

Whole-brain calcium imaging by two-photon microscopy

To monitor neural activity throughout the zebrafish brain, we used Tol2-mediated transgenesis to generate a stable line expressing GCaMP5G[25] under the control of the HuC promoter. To noninvasively image throughout the brain, we constructed a custom two-photon microscope. Microscope hardware was controlled by custom software written in C# (Microsoft), except photon counting, which was implemented in C++ to

perform peak detection on PMT signals digitized at 200 MHz (Spectrum). Upon acquiring an image from the behavioral camera, we recorded the index of the image currently being scanned to allow synchronization of calcium activity and behavioral events.

Motion-stabilized calcium imaging during behavior

Our imaging strategy was based on selecting a single focal plane and then stably monitoring the same focal plane for the duration of the behavioral paradigm, roughly two hours. To mitigate the effects of motor-induced and thermal stimulus-induced motion artifacts, we performed online focal plane correction and offline lateral correction. Prior to starting the behavioral paradigm, we first collected an anatomical reference stack 40 µm above and below the chosen focal plane. We then collected images continuously throughout the experimental paradigm. Each newly scanned image was cross correlated in three dimensions with the reference stack using Intel Performance Primitives IPP) and C#. The lateral search size was +/-10 µm and the axial search size was +/-5 µm (11 image slices of the reference stack). The depth of the best z slice was low pass filtered and kept within 1 µm of the original focal plane by adjusting the objective height in 1 µm increments. Lateral image registration was performed offline using Matlab and IPP to fully take advantage of the entire time series.

Image analysis pipeline

After offline lateral image registration, we calculated a morphology image by averaging each pixel across the entire time series. We exploited the fact that GCaMP5 is excluded from the nucleus to identify cell centroids by searching for local minima that were also

below an absolute intensity threshold. We defined the final mask of pixels belonging to the cell based on the intersection of a morphologically defined mask and a functionally defined mask. To obtain the functionally defined mask, we first obtained a preliminary activity trace using a circular mask with a conservatively chosen radius of 1 μm. For each neighboring pixel up to a radius of 4.5 μm, we calculated its correlation coefficient with this preliminary activity trace. We obtained the final mask of pixels belonging to the cell by adaptively lowering a threshold until the resulting thresholded mask covered at least 30 μm^2. The final mask, obtained by intersecting the morphologically and functionally defined masks, was used to extract a final activity trace for each neuron. To isolate heat on, heat on, ITI, motor components of each active trace, we perform linear unmixing using the impulse response function of each component as the kernel for linear regression.

Circuit trajectories

To analyze changes in circuit state across trials, we performed principal component analysis (PCA) in Matlab. In our approach, we used the trials of both blocks as the observations. For the variables, we used the activity of every neuron in a given brain region at every time point within the trial. To visualize the circuit trajectories of individual brain regions across trials, we projected the population activity data of each brain region onto the first three principal components.

Response classification

For all stimulus related classes of neurons, we first filter out cells with non-specific activity, where ITI activity accounts for most of variance in the data. A neuron is

classified as a motor neuron if most of the variance in its activity is accounted for by the motor related signals. We remove motor-related neurons from classification of sensory and learning related classes. Sensory neurons are defined as neurons with < 0.1 correlation with past performance, and differential activity between error and correct trials of < 0.5. RP+, RP- neurons are neurons with > 0.2 and <-0.2 correlation with past performance. RPE neurons are neurons with > 0 correlation with past performance for the heat on component of its activity and < 0 correlation with past performance for the heat off component of its activity. We additionally constrain the classification by setting limits on the ratio of correct trial to error trial peak activity for each given class. RP+ and RPE neurons were constrained to a ratio >0.9, and RP- neurons were constrained to a ratio of < 1 for the heat on component of the activity. Turning-related neurons and action selection neurons are identified by ROC analysis. Action selection neurons were required to be differentially active before heat-evoked and spontaneous turns, as well as differentially active before left or right heat-evoked turns.

Chapter Four

Discussions and future directions

Abstract

We have developed the first zebrafish operant learning paradigm and shown that larval zebrafish are capable of rapid and stable operant learning. In this section, we will place our paradigm in the larger context of the many operant learning paradigms that have been established in the mammalian literature. We will discuss the interpretation of learning impairments that we observe across the larval population and consider the possible advantages of exploiting natural learning deficits instead of artificially inducing them. We will also discuss possible future directions for our behavioral work.

We have also performed the first functional analysis of learning circuitry across the larval zebrafish brain. We have analyzed the types of signals present during learning, as well as the distribution of these signals across the brain and across the left-right axis. We have specifically uncovered a functionally asymmetric habenula circuit that may encode relief prediction and relief prediction error, as well as neuronal activity that anticipates the direction of heat evokes turns. We will discuss possible future directions to more fully characterize these learning related signals as well as variations to the learning paradigm that will improve our ability to dissect the learning circuitry.

Conclusions and discussion

We demonstrate that larval zebrafish are capable of operant learning. Our learning paradigm is able to quickly and reliably entrain a turning bias in response to heat onset. This bias is stimulus dependent and retained for at least one hour. This is the first report of operant learning in larval zebrafish, and one of only two reports of associative learning at the larval stage. Though larval zebrafish have a uniquely accessible brain

and many available genetic tools, the general lack of robust associative learning paradigms in larval zebrafish has hindered investigations of learning in this animal. In contrast, the long history and proliferation of increasingly sophisticated learning paradigms in mammals have yielded a number of important insights into operant learning behavior and the neural circuitry that underlies learning. Our learning paradigm could serve as an entry point for studying learning behavior and circuitry in larval zebrafish,

Operant learning behavior

How does our learning paradigm compare with operant learning in mammalian systems?

Four canonical categories of operant learning have being defined for mammals[2]. These paradigms can be divided along two axes – whether the reinforcer is appetitive or aversive, and whether the response promotes or suppresses the reinforcer. Our learning paradigm fits well into the category of escape learning, in which the correct response (e.g. left turn) can terminate an aversive stimulus (e.g. noxious heat).

Beyond these four basic categories, a diverse array of more sophisticated operant learning paradigms has emerged over the years in various mammalian systems. One of the most significant developments in our current understanding of operant learning is the idea that the R-O association is not the only association that can be formed during operant learning[2]. Another association can be formed between contextual sensory cues and the response. In other words, the response is context specific. For example, a pigeon trained to peck a red dot for food, does not begin to peck randomly at all objects.

The pecking response is modulated by the presence of the red dot. The actual relationship learned during operant learning is S →(R-O). In our learning paradigm, S is heat onset, R is directional turning, and O is the laser state after the first turn. Consistent with most operant learning paradigms in mammals, we find that performance of R is dependent on the presence of S, while spontaneous turns are unbiased.

More recent studies of operant learning have introduced a further distinction in operant learning between goal directed learning and habit learning. The distinction rests on which relationships are formed between S, R. and O during training and performance of the behavior. In goal directed behavior, it is hypothesized that the animals internally represent the desired outcome and its relationship to the response (O ←→ R). The overall relationship in a goal directed behavior can be represented as S → (O ←→ R).

After extensive training, numerous experiments have shown that the intermediate state in which the outcome is internally represented is no longer accessible. The animal now operates out of habit, and the stimulus directly activates the learned response, S → R. Outcome devaluation and contingency degradation experiments have demonstrated the existence of the O → R relationship, and verified that this relationship no longer governs the behavior after overtraining. In our learning paradigm, we have demonstrated that after 25-30 trials, the response can be extinguished in open loop experiments as well as reversed in a reversal training block. This suggests that within the time frame of our experiment, the behavior is still dependent on R-O relationships.

Population analysis of learning behavior

Though more complex variations of our learning paradigm are needed to assay the learning process in greater depth, there is an advantage to the simplicity and speed of our assay. For example, we have assayed learning in over 200 animals. At a population level, we can ask not only questions about the speed of acquisition or the duration of retention, but also whether there are systematic patterns to learning behavior or ability.

We find that zebrafish exhibit two specific learning impairments that occur independently – switching impairment and retention impairment. Switching-impaired animals strongly prefer to turn in one direction in response to heat. These animals switch turn direction at a much lower rate and often fail to terminate the heat during error trials. Spontaneous turning is generally unaffected, and the animal is capable of turning in both directions in response to heat onset.

We also find animals that are retention-impaired. These animals have normal switching behavior, and can successfully terminate all error trials. However, these animals do not retain a preference for the rewarded turn. The asymptotic performance of these animals is on average around 50%, indicating a lack of preference for either turn direction. Furthermore, we can show that this impairment is dependent on developmental stage. Almost all animals exhibit impaired retention at 3-4 dpf, while less than a quarter of animals exhibit impaired retention at 6-7 dpf.

We believe that there is a distinct advantage to looking at natural variation in learning ability and its development over time. Our data indicates that the brain appears to have two stereotyped failure modes. How does this happen? Is there a reason why some circuits are more susceptible to disruption than others? Which factors determine why

some circuits are more prone to developmental delays or mis-wiring? Mutant and drug analysis can create an artificial problem that the brain is not designed to solve. Analyzing naturally occurring impairments can potentially yield insights into the robustness of different circuit components and the question of why some neurological disorders are more prevalent than others.

Future Directions

There are several extensions of our experiments that could yield valuable insight into the learning process.

1) How specific is the stimulus in our learning paradigm? We were able to show that the learned response was context dependent. Spontaneous turns were unbiased after training, while heat driven turns were biased in the rewarded turn direction. However, our experiments do not distinguish whether the bias is associated specifically with noxious heat or with any noxious sensory stimulus. To distinguish between these possibilities, we will train animals with the heat stimulus and then test turn direction bias in response to other noxious stimuli (i.e. chemical stimuli such as mustard oil).
2) Is larval zebrafish capable of goal directed learning? Our current experiments cannot conclusively distinguish between habit learning and goal directed learning. Outcome devaluation experiments are most amenable to studies that involve some form of appetitive conditioning. However, we can assay whether extinction becomes progressively more difficult as training progresses and whether a point can be reached when the animal is no longer sensitive to any changes in outcome. This

would not definitively show that the larval zebrafish has an internal representation of outcome, but would determine whether the learning becomes increasingly habitual (S → R) as training progresses.

3) We were able to use noxious heat as a reinforcement signal in larval zebrafish. Previous efforts within the lab using electrical shock and touch have failed to consistently reinforce learning behavior. However, these experiments used a different set of parameters and are not directly comparable to our paradigm. We will address whether at 7 dpf, larval zebrafish can only use noxious heat as a reinforcement signal or whether other stimuli can be used in operant learning. To this end, we will attempt to train larvae with other noxious stimuli (i.e. chemical) or appetitive stimuli (i.e. food odorant).
4) We have some evidence that animals can be trained to reverse their motor response over multiple training blocks. The difficulty in extending training past two blocks is the increased fraction of animals that experience physical damage and undergo seizure events. However, in the fraction of animals that can be trained to undergo multiple reversal blocks, one can ask whether the process of meta-learning can be observed. We would do this by alternating the rewarded turn direction for each block. To control for innate biases in turn direction, we would then compare performance across the blocks that reward the same turn direction, and determine whether the rate of learning changes over successive blocks.

Learning Circuitry

Our imaging and analysis of neural activity across the entire larval brain has yielded insights both into the types of signals present during learning but also the relative distribution of these signals across brain regions and across the left-right axis.

Learning related signals in the habenula

We find most learning related signals to be concentrated in the forebrain including the habenula, pallium, and subpallium, while motor signals dominate the hindbrain. This is consistent with the evolutionary view that the pallium is homologous to many of the learning-related structures in mammals (i.e. amygdala, hippocampus, and striatum), and the habenula is conserved across most vertebrates and mediates aversive learning. The lateral habenula (LHb) in particular has been implicated in avoidance learning in both mammals and fish[47-67, 87]. LHb is believed to encode negative motivational value and interact closely with the midbrain dopamine system to form the reward prediction error signal[47, 60-67]. Our analysis suggests that the habenula contains not only negative prediction error, but also signals that encode relief prediction error, as well as heat onset in the presence or absence of relief prediction. Our results suggests that 1) a noxious events such as heat onset can act as a cue for relief prediction, 2) the habenula contains a broader range of prediction and prediction error signals than previously believed, and 3) the prediction-related signals in the habenula are asymmetrically distributed across the left and right axis.

Anatomical and genetically specified lateralization of the habenula is well known in both mammals and fish. In zebrafish, the left habenula is specified by nodal signaling during

early development, and can be differentiated from the right habenula both by anatomy and connectivity[130-133]. The right habenula receives inputs from left and right pallium, while the left habenula receives inputs from the left pallium. Both sides of the habenula receive inputs from the eminentia thalamus (EMT). The right habenula projects to the ventral IPN and raphe, while the left habenula projects to the dorsal IPN and raphe. The left habenula is also 20% larger than the right habenula by adulthood.

Surprisingly, little is known about functional lateralization in the habenula. Very few studies have shown that any behavior can be correlated with habenula lateralization[133]. Furthermore, the intermediate and crucially important step of characterizing which functional signals are lateralized in the habenula is nonexistent. Our analysis demonstrates that prediction related clusters are asymmetrically distributed across the left and right habenula. In contrast, functional modules in the hindbrain, midbrain, pallium, and subpallium tend to be symmetric across the left-right axis. It remains to be seen whether the distribution of these relief prediction signals in the habenula is specific to zebrafish, or is conserved across vertebrates.

Future Directions

1) We characterize a number of relief prediction related neuron in the forebrain. The presence of these signals are consistent with the interpretation that heat onset can act as a cue/predictor for relief. Though canonical learning paradigms generally use a neutral cue as a predictor for reward, punishment, or relief, there is no reason to believe that a noxious event cannot also act as a predictor. To determine whether the relief prediction signals we find are particular to this learning paradigm, we will implement a

variant of our learning paradigm to include neutral predictive cues. In this version, the duration of heat onset is randomized per trial, but a neutral cue precedes heat offset with a predictable delay.

4) For canonical reward based learning paradigms, the temporal difference (TD) model suggests that inhibitory connections must exist between reward prediction and reward prediction error neurons. It is unclear whether the same model applies to the relief prediction signals we find. We can take several approaches to determine whether the RPE neurons are inhibited by RP+ neurons in the habenula. At the most basic level, we will use co-expression analysis to determine the neurotransmitters type for each relief prediction signal. If the interaction between RPE and RP+ follow the TD model, then we expect to find at least a subset of RP+ neurons to be inhibitory. To most rigorously address the more difficult question of connectivity, we would need to use dual patching electrophysiology, viral tracing, or EM reconstruction. None of these technologies fully established for larval zebrafish, and not currently feasible for addressing the connectivity between relief prediction signals in the habenula. Targeted Channelrhodopsin excitation may be also allow us to address connectivity, though this method cannot distinguish between monosynaptic and polysynaptic connections without simultaneous electrophysiological recordings. However, single cell targeting of functionally identified cells by ChR2 has yet to be done in larval zebrafish.

3) We find block-specific activity in a number of neurons. A subset of these neurons may mediate action selection while another subset may encode response outcome. However, it is an open question whether the same set of neurons are activated during all training blocks with the same reward contingency. Specifically, if we alternate reward

contingency in each training block, is the network activity in the first block the same as the network activity in the third block? If the same stereotyped network is activated whenever animals learn to turn in a particular direction, then we would expect the same network activity in the first and third blocks. However, though the reward contingency is the same in these two blocks, they are also different in terms of learning history. In the first block, the brain must transition from a naïve state, while in the third block, the brain must inhibit a previously learned response and switch its behavioral output. This difference may be reflected in the network activity. It would be interesting to image the brain through three alternating training blocks to distinguish between these two possibilities.

However, there are some definite limitations to extending the number of training blocks. In the small subset of animals that we have imaged for more than two training blocks, there are higher instances of physical damage and seizure-like events in the brain. These are generally not observed through the first two training blocks. If we were to image more than two training blocks, animals that show seizure-like events and animals that fail to learn in the third training block would have to be excluded.

4) As we have mentioned in the previous chapter, it would be interesting to train with other noxious stimuli to determine whether the learning-related signals we observe are specific to the sensory modality of the reinforcement, or if the same network is activated by all noxious stimuli. We know based on imaging in the peripheral sensory ganglion that different subsets of sensory cells mediate different sensory modalities (i.e. noxious heat is distinct from harsh touch). However, do all noxious stimulus modalities converge in the downstream circuitry to form a generalized pain signal that then interacts with

error signals in the habenula and dopaminergic system, or is each stimulus modality separately encoded even in the habenula? For example, are the neurons activated by heat in error trials generalized error signals across all modalities or are they modality specific? To address this question we must reinforce animals with two different sensory modalities, and determine which active neurons are common to both modalities and which are distinct.

5) We would ideally like to more finely map the network activity across brain regions. Using anatomical landmarks, the resolution is limited. Prominent structures such as habenula and cerebellum can be easily defined, pallium can be separated from the subpallium, and the boundary between midbrain and hindbrain is clear. However, there are many structures that may require co-labeling with genetic markers to visualize. For example, the lateral and medial habenula are not distinguishable anatomically in larvae, but there are transgenic lines that specifically label the lateral habenula. The same is true for the catecholinergic system. Dopaminergic neurons reside in the ventral midbrain and are difficult to isolate anatomically, but there are now several transgenic lines that label at least a subset of the dopaminergic neurons.

References

1. Pearce JM, Bouton ME. Theories of associative learning in animals. Annu Rev Psychol. 2001;52:111-39.
2. Bouton, ME. Learning and behavior: A contemporary synthesis. (2007) Sunderland, MA. Sinauer. Pp. vii + 482. ISBN-13: 978-0-87893-063-0
3. Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasy WF, editors. Classical Conditioning II: Current Research and Theory. New York, New York: Appleton Century Crofts; 1972. pp. 64–99.
4. Pavlov I. P. Conditioned reflexes. (2007) London: Oxford University Press
5. Thorndike E. L. Animal intelligence: experimental studies. (1911) NewYork:Macmillan
6. Rescorla RA, Solomon RL. Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. Psychol Rev. 1967 May;74(3):151-82.
7. Balleine BW, Ostlund SB. Still at the choice point: Action selection and initiation in instrumental conditioning. Ann NY Acad Sci. 2007d;1104:147–171.
8. Dickinson A., Balleine B., Watt A., Gozalez F., Boakes R.A. Motivational control after extended instrumental learning. (1995) Animal Learning and Behavior. 23, 197-206
9. Colwill R.M., Rescorla R.A. Effect of reinforce devaluation on discriminative control of instrumental behavior. Journal of Experimental Psychology: Animal Behavior Processes. (1990) 16, 40-47
10. Holland P.C. Amount of training effects in representation-mediated food aversion learning: no evidence of a role for associability changes. Learn Behav. 2005 Nov;33(4):464-78.
11. Holland PC. Relations between Pavlovian-instrumental transfer and reinforcer devaluation. J Exp Psychol Anim Behav Process. 2004 Apr;30(2):104-17.
12. Balleine BW, Dickinson A. The role of incentive learning in instrumental outcome revaluation by specific satiety. Anim Learn Behav. 1998b;26:46–59.
13. Papini MR. Comparative Psychology. (2002) Upper Saddle River, NJ: Prentice Hall
14. Hikosaka O, Isoda M. Switching from automatic to controlled behavior: cortico-basal ganglia mechanisms. Trends Cogn Sci. 2010 Apr;14(4):154-61. Epub 2010 Feb 22.
15. Isoda M, Hikosaka O. Cortico-basal ganglia mechanisms for overcoming innate, habitual and motivational behaviors. Eur J Neurosci. 2011 Jun;33(11):2058-69.
16. Johnston K, Levin HM, Koval MJ, Everling S. Top-down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. Neuron. 2007 Feb 1;53(3):453-62.
17. Schultz, W. & Dickinson, A. Neuronal coding of prediction errors. Annu. Rev. Neurosci. 23, 473–500 (2000).
18. Bayer HM, Glimcher PW. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron. 2005;47:129–141.

19. Schultz W. Predictive reward signal of dopamine neurons. J Neurophysiol. 1998;80:1–27.
20. Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science.1997;275:1593–1599. [PubMed]
21. Montague PR, Dayan P, Sejnowski TJ (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 16: 1936–1947.
22. Schultz W. Behavioral theories and the neurophysiology of reward. Annu Rev Psychol. 2006;57:87-115.
23. Hikosaka, O. Basal ganglia mechanisms of reward oriented eye movement. Ann. N. Y. Acad. Sci. 1104, 229–249 (2007).
24. Yin HH, Ostlund SB, Knowlton BJ, Balleine BW. 2005. The role of the dorsomedial striatum in instrumental conditioning Eur J Neurosci 22513–523.523
25. Lauwereyns J, Watanabe K, Coe B, Hikosaka O. A neural correlate of response bias in monkey caudate nucleus. Nature. 2002 Jul 25;418(6896):413-7.
26. Samejima K, Ueda Y, Doya K, Kimura M. Representation of action-specific reward values in the striatum. Science. 2005 Nov 25;310(5752):1337-40.
27. Watanabe K, Lauwereyns J, Hikosaka O. Neural correlates of rewarded and unrewarded eye movements in the primate caudate nucleus. J Neurosci 2003;23:10052–10057
28. Hikosaka O, Sakamoto M, Usui S. Functional properties of monkey caudate neurons. I activities related to saccadic eye movements. J Neurophysiol 1989a;61:780–798.
29. Hikosaka O, Sakamoto M, Usui S. Functional properties of monkey caudate neurons. III activities related to expectation of target and reward. J Neurophysiol 989b;61:814–832.
30. Balleine BW, Delgado MR, Hikosaka O. The role of the dorsal striatum in reward and decision-making. J Neurosci. 2007a;27:8161–8165.
31. Ostlund SB, Balleine BW. The contribution of orbitofrontal cortex to action selection. Ann N Y Acad Sci. 2007 Dec;1121:174-92. Epub 2007 Sep 13.
32. Yin HH, Knowlton BJ, Balleine BW. 2004. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning Eur J Neurosci 19181–189.189
33. Yin HH, Knowlton BJ, Balleine BW. Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. Behav Brain Res. 2006;166:189–196.
34. Balleine BW, Dickinson A. 1998a. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates Neuropharmacology 37407–419.419
35. Balleine BW, O'Doherty JP. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology. 2010 Jan;35(1):48-69.
36. Bromberg-Martin ES, Matsumoto M, Hikosaka O. Dopamine in motivational control: rewarding, aversive, and alerting. Neuron. 2010 Dec 9;68(5):815-34.
37. Schultz W. Multiple dopamine functions at different time courses. Annual review of neuroscience.2007;30:259–288.

38. Schultz W. Dopamine signals for reward value and risk: basic and recent data. Behav Brain Funct.2010;6:24. [PMC free article] [PubMed]
39. Matsumoto M, Hikosaka O. Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature. 2009b;459:837–841.
40. Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature. 2012 Jan 18;482(7383):85-8.
41. Joshua M, Adler A, Mitelman R, Vaadia E, Bergman H. Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probabilistic classical conditioning trials. J Neurosci. 2008;28:11673–11684.
42. Tobler PN, Fiorillo CD, Schultz W. Adaptive coding of reward value by dopamine neurons. Science.2005;307:1642–1645.
43. Horvitz JC. Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events.Neuroscience. 2000;96:651–656.
44. Bromberg-Martin ES, Hikosaka O. Midbrain dopamine neurons signal preference for advance information about upcoming rewards. Neuron. 2009;63:119–126.
45. Bromberg-Martin ES, Matsumoto M, Hikosaka O. Distinct tonic and phasic anticipatory activity in lateral habenula and dopamine neurons. Neuron. 2010a;67:144–155.
46. Bromberg-Martin ES, Matsumoto M, Nakahara H, Hikosaka O. Multiple timescales of memory in lateral habenula and dopamine neurons. Neuron. 2010c;67:499–510.
47. Hikosaka O. The habenula: from stress evasion to value-based decision-making. Nat Rev Neurosci. 2010 Jul;11(7):503-13.
48. Concha, M. L. & Wilson, S. W. Asymmetry in the epithalamus of vertebrates. J. Anat. 199, 63–84 (2001).
49. Herkenham, M. & Nauta, W. J. Afferent connections of the habenular nuclei in the rat. A horseradish peroxidase study, with a note on the fiber-of-passage problem. J. Comp. Neurol. 173, 123–146 (1977).
50. Herkenham, M. & Nauta, W. J. Efferent connections of the habenular nuclei in the rat. J. Comp. Neurol. 187, 19–47 (1979).
51. Jhou, T. C., Geisler, S., Marinelli, M., Degarmo, B. A. & Zahm, D. S. The mesopontine rostromedial tegmental nucleus: a structure targeted by the lateral habenula that projects to the ventral tegmental area of Tsai and substantia nigra compacta. J. Comp. Neurol. 513, 566–596 (2009).
52. Hong, S. & Hikosaka, O. The globus pallidus sends reward-related signals to the lateral habenula. Neuron 60, 720–729 (2008).
53. Dafny, N. & Qiao, J. T. Habenular neuron responses to noxious input are modified by dorsal raphe stimulation. Neurol. Res. 12, 117–121 (1990).
54. Kim, U. Topographic commissural and descending projections of the habenula in the rat. J. Comp. Neurol. 513, 173–187 (2009).
55. Jhou, T. C., Fields, H. L., Baxter, M. G., Saper, C. B. & Holland, P. C. The rostromedial tegmental nucleus (RMTg), a GABAergic afferent to midbrain dopamine neurons, encodes aversive stimuli and inhibits motor responses. Neuron 61, 786–800 (2009).

56. Aghajanian, G. K. & Wang, R. Y. Habenular and other midbrain raphe afferents demonstrated by a modified retrograde tracing technique. Brain Res. 122, 229–242 (1977).
57. Stern, W. C., Johnson, A., Bronzino, J. D. & Morgane, P. J. Effects of electrical stimulation of the lateral habenula on single-unit activity of raphe neurons. Exp. Neurol. 65, 326–342 (1979).
58. Nishikawa, T. & Scatton, B. Inhibitory influence of GABA on central serotonergic transmission. Involvement of the habenulo-raphe pathways in the GABAergic inhibition of ascending cerebral serotonergic neurons. Brain Res. 331, 81–90 (1985).
59. Lecourtier, L. & Kelly, P. H. A conductor hidden in the orchestra? Role of the habenular complex in monoamine transmission and cognition. Neurosci. Biobehav. Rev. 31, 658–672 (2007).
60. Matsumoto, M. & Hikosaka, O. Lateral habenula as a source of negative reward signals in dopamine neurons. Nature 447, 1111–1115 (2007).
61. Matsumoto, M. & Hikosaka, O. Representation of negative motivational value in the primate lateral habenula. Nature Neurosci. 12, 77–84 (2009).
62. Ji, H. & Shepard, P. D. Lateral habenula stimulation inhibits rat midbrain dopamine neurons through a GABA(A) receptor-mediated mechanism. J. Neurosci. 27, 6923–6930 (2007).
63. Nishikawa, T., Fage, D. & Scatton, B. Evidence for, and nature of, the tonic inhibitory influence of habenulointerpeduncular pathways upon cerebral dopaminergic transmission in the rat. Brain Res. 373, 324–336 (1986).
64. Ullsperger, M. & von Cramon, D. Y. Error monitoring using external feedback: specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. J. Neurosci. 23, 4308–4314 (2003).
65. Christoph, G. R., Leonzio, R. J. & Wilcox, K. S. Stimulation of the lateral habenula inhibits dopamine containing neurons in the substantia nigra and ventral tegmental area of the rat. J. Neurosci. 6, 613–619 (1986).
66. Rausch, L. J. & Long, C. J. Habenular Lesions and avoidance learning deficits in allbins rats. Physiol. Behav. 2, 352–356 (1974).
67. Matsumoto M, Hikosaka O. Electrical stimulation of the primate lateral habenula suppresses saccadic eye movement through a learning mechanism. PLoS One. 2011;6(10):e26701. Epub 2011 Oct 24.
68. Wullimann MF, Rink E. The teleostean forebrain: a comparative and developmental view based on early proliferation, Pax6 activity and catecholaminergic organization. Brain Res Bull. 2002 Feb-Mar 1;57(3-4):363-70.
69. Smeets WJ. Comparative aspects of basal forebrain organization in vertebrates. Eur J Morphol. 1992;30(1):23-36.
70. Butler AB, Hodos H. Comparative vertebrate neuroanatomy: evolution and adaptation. Wiley-Liss, New York, 1996
71. Bianco IH, Wilson SW. The habenular nuclei: a conserved asymmetric relay station in the vertebrate brain. Philos Trans R Soc Lond B Biol Sci. 2009 Apr 12;364(1519):1005-20.

72. Okamoto H, Agetsuma M, Aizawa H. Genetic dissection of the zebrafish habenula, a possible switching board for selection of behavioral strategy to cope with fear and anxiety. Dev Neurobiol. 2012 Mar;72(3):386-94. doi: 10.1002/dneu.20913.
73. Xi Y, Yu M, Godoy R, Hatch G, Poitras L, Ekker M. Transgenic zebrafish expressing green fluorescent protein in dopaminergic neurons of the ventral diencephalon. Dev Dyn. 2011 Nov;240(11):2539-47. doi: 10.1002/dvdy.22742
74. Kastenhuber E, Kratochwil CF, Ryu S, Schweitzer J, Driever W. Genetic dissection of dopaminergic and noradrenergic contributions to catecholaminergic tracts in early larval zebrafish. J Comp Neurol. 2010 Feb 15;518(4):439-58.
75. Lillesaar C, Stigloher C, Tannhäuser B, Wullimann MF, Bally-Cuif L. Axonal projections originating from raphe serotonergic neurons in the developing and adult zebrafish, Danio rerio, using transgenics to visualize raphe-specific pet1 expression. J Comp Neurol. 2009 Jan 10;512(2):158-82.
76. Ma PM. Catecholaminergic systems in the zebrafish. II. Projection pathways and pattern of termination of the locus coeruleus. J Comp Neurol. 1994 Jun 8;344(2):256-69.
77. Ma PM. Catecholaminergic systems in the zebrafish. I. Number, morphology, and histochemical characteristics of neurons in the locus coeruleus. J Comp Neurol. 1994 Jun 8;344(2):242-55.
78. Hashimoto M, Hibi M. Development and evolution of cerebellar neural circuits. Dev Growth Differ. 2012 Apr;54(3):373-89.
79. Salas C, Broglio C, Durán E, Gómez A, Ocaña FM, Jiménez-Moya F, Rodríguez F. Neuropsychology of learning and memory in teleost fish. Zebrafish. 2006;3(2):157-71.
80. Rodríguez F, López JC, Vargas JP, Gómez Y, Broglio C, Salas C. Conservation of spatial memory function in the pallial forebrain of reptiles and ray-finned fishes. J Neurosci. 2002 Apr 1;22(7):2894-903.
81. López JC, Bingman VP, Rodríguez F, Gómez Y, Salas C. Dissociation of place and cue learning by telencephalic ablation in goldfish. Behav Neurosci. 2000 Aug;114(4):687-99.
82. Portavella M, Torres B, Salas C. Avoidance response in goldfish: emotional and temporal involvement of medial and lateral telencephalic pallium. J Neurosci. 2004 Mar 3;24(9):2335-42.
83. Martín I, Gómez A, Salas C, Puerto A, Rodríguez F. Dorsomedial pallium lesions impair taste aversion learning in goldfish. Neurobiol Learn Mem. 2011 Sep;96(2):297-305. Epub 2011 Jun 13.
84. Gómez A, Durán E, Salas C, Rodríguez F. Cerebellum lesion impairs eyeblink-like classical conditioning in goldfish. Neuroscience. 2010 Mar 10;166(1):49-60. Epub 2009 Dec 16.
85. Romaguera F, Mattioli R. Chlorpheniramine impairs spatial choice learning in telencephalon-ablated fish. Biol Res. 2008;41(3):341-8. Epub 2009 Apr 23.
86. Rodríguez F, Durán E, Gómez A, Ocaña FM, Alvarez E, Jiménez-Moya F, Broglio C, Salas C. Cognitive and emotional functions of the teleost fish cerebellum. Brain Res Bull. 2005 Sep 15;66(4-6):365-70. Epub 2005 Feb 25.
87. Agetsuma M, Aizawa H, Aoki T, Nakayama R, Takahoko M, Goto M, Sassa T, Amo R, Shiraki T, Kawakami K, Hosoya T, Higashijima S, Okamoto H. The habenula is

crucial for experience-dependent modification of fear responses in zebrafish. Nat Neurosci. 2010 Nov;13(11):1354-6. Epub 2010 Oct 10.
88. Woodard,W. T., and Bitterman, M. E. (1973). Pavlovian analysis of avoidance conditioning in the goldfish (Carassius auratus). J. Comp. Physiol. Psychol. 82, 123–129.
89. Sison, M., and Gerlai, R. (2010). Associative learning in zebrafish (Danio rerio) in the plus maze. Behav. Brain Res. 207, 99–104.
90. Pather, S., and Gerlai, R. (2009). Shuttle box learning in zebrafish. Behav. Brain Res. 196, 323–327.
91. Xu, X., Scott-Scheiern, T., Kempker, L., and Simons, K. (2007). Active avoidance conditioning in zebrafish (Danio rerio). Neurobiol. Learn. Mem. 87, 72–77.
92. Gomez-Laplaza, L. M., and Gerlai, R. (2010). Latent learning in zebrafish (Danio rerio). Behav. Brain Res. 208, 509–515
93. Braubach, O. R.,Wood, H. D., Gadbois, S., Fine, A., and Croll, R. P. (2009). Olfactory conditioning in the zebrafish (Danio rerio). Behav. Brain. Res. 198, 190–198
94. Best, J. D., Berghmans, S., Hunt, J. J., Clarke, S. C., Fleming, A., Goldsmith, P., Roach, A. G. (2008). Nonassociative learning in larval zebrafish. Neuropsychopharmacology 33, 1206–1215.
95. Aizenberg M, Schuman EM. Cerebellar-dependent learning in larval zebrafish. J Neurosci. 2011 Jun 15;31(24):8708-12.
96. Valente A, Huang KH, Portugues R, Engert F. Ontogeny of classical and operant learning behaviors in zebrafish. Learn Mem. 2012 Mar 20;19(4):170-7.

97. Tian L, Hires SA, Mao T, Huber D, Chiappe ME, Chalasani SH, et al. Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. Nat Methods. 2009;6(12):875-81.
98. Miyawaki A, Llopis J, Heim R, McCaffery JM, Adams JA, Ikura M, Tsien RY. Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. Nature. 1997 Aug 28;388(6645):882-7.
99. Mank M, Santos AF, Direnberger S, Mrsic-Flogel TD, Hofer SB, Stein V, Hendel T, Reiff DF, Levelt C, Borst A, Bonhoeffer T, Hübener M, Griesbeck O. A genetically encoded calcium indicator for chronic in vivo two-photon imaging. Nat Methods. 2008 Sep;5(9):805-11.
100. Nagai T, Sawano A, Park ES, Miyawaki A. Circularly permuted green fluorescent proteins engineered to sense Ca2+. Proc Natl Acad Sci U S A. 2001 Mar 13;98(6):3197-202.
101. Akerboom J, Chen TW, Wardill TJ, Tian L, Marvin JS, Mutlu S, Calderón NC, Esposti F, Borghuis BG, Sun XR, Gordus A, Orger MB, Portugues R, Engert F, Macklin JJ, Filosa A, Aggarwal A, Kerr RA, Takagi R, Kracun S, Shigetomi E, Khakh BS, Baier H, Lagnado L, Wang SS, Bargmann CI, Kimmel BE, Jayaraman V, Svoboda K, Kim DS, Schreiter ER, Looger LL. Optimization of a GCaMP Calcium Indicator for Neural Activity Imaging. J Neurosci. 2012 Oct 3;32(40):13819-13840.
102. Huber D, Gutnisky DA, Peron S, O'Connor DH, Wiegert JS, Tian L, Oertner TG, Looger LL, Svoboda K. Multiple dynamic representations in the motor cortex during sensorimotor learning. Nature. 2012 Apr 25;484(7395):473-8.

103. Komiyama T, Sato TR, O'Connor DH, Zhang YX, Huber D, Hooks BM, Gabitto M, Svoboda K. Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. Nature. 2010 Apr 22;464(7292):1182-6.
104. Harvey CD, Coen P, Tank DW. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature. 2012 Mar 14;484(7392):62-8.
105. Dombeck DA, Harvey CD, Tian L, Looger LL, Tank DW. Functional imaging of hippocampal place cells at cellular resolution during virtual navigation. Nat Neurosci. 2010 Nov;13(11):1433-40. Epub 2010 Oct 3.
106. Seelig JD, Chiappe ME, Lott GK, Dutta A, Osborne JE, Reiser MB, Jayaraman V. Two-photon calcium imaging from head-fixed Drosophila during optomotor walking behavior. Nat Methods. 2010 Jul;7(7):535-40. Epub 2010 Jun 6.
107. Wolf R, Heisenberg M. Visual control of straight flight in Drosophila melanogaster. J Comp Physiol A. 1990 Jul;167(2):269-83.
108. Brembs B, Heisenberg M. The operant and the classical in conditioned orientation of Drosophila melanogaster at the flight simulator. Learn Mem. 2000 Mar-Apr;7(2):104-15.
109. Ahrens MB, Li JM, Orger MB, Robson DN, Schier AF, Engert F, Portugues R. Brain-wide neuronal dynamics during motor adaptation in zebrafish. Nature. 2012 May 9;485(7399):471-7.
110. White NM, McDonald RJ. Multiple parallel memory systems in the brain of the rat. Neurobiol Learn Mem 2002 77:125–184.
111. Poldrack RA, Packard MG. Competition among multiple memory systems: converging evidence from animal and human brain studies. Neuropsychologia 2003 41:245–251
112. Karlsson MP, Tervo DG, Karpova AY. Network resets in medial prefrontal cortex mark the onset of behavioral uncertainty. Science. 2012 Oct 5;338(6103):135-9.
113. R. Quilodran, M. Rothé, E. Procyk, Behavioral shifts and action valuation in the anterior cingulate cortex. Neuron 57, 314 (2008). CrossRefMedlineWeb of Science
114. D. Durstewitz, N. M. Vittoz, S. B. Floresco, J. K. Seamans, Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. Neuron 66, 438 (2010). CrossRefMedlineWeb of Science
115. Buonomano DV, Maass W. State-dependent computations: spatiotemporal processing in cortical networks. Nat Rev Neurosci. 2009 Feb;10(2):113-25.
116. Stopfer M, Jayaraman V, Laurent G. Intensity versus identity coding in an olfactory system. Neuron. 2003 Sep 11;39(6):991-1004.
117. Niessing J, Friedrich RW. Olfactory pattern classification by discrete neuronal network states. Nature. 2010 May 6;465(7294):47-52.
118. Dombeck DA, Graziano MS, Tank DW. Functional clustering of neurons in motor cortex determined by cellular resolution imaging in awake behaving mice. J Neurosci. 2009 Nov 4;29(44):13751-60.
119. Gemberling GA, Domjan M. Selective associations in one-day-old rats: taste-toxicosis and texture-shock aversion learning. J Comp Physiol Psychol. 1982 Feb;96(1):105-13.
120. Johanson IB, Hall WG. Appetitive learning in 1-day-old rat pups. Science. 1979 Jul 27;205(4404):419-21.

121. Dehorter N, Vinay L, Hammond C, Ben-Ari Y. Timing of developmental sequences in different brain structures: physiological and pathological implications. Eur J Neurosci. 2012 Jun;35(12):1846-56.
122. Antonopoulos J, Dori I, Dinopoulos A, Chiotelli M, Parnavelas JG. Postnatal development of the dopaminergic system of the striatum in the rat. Neuroscience. 2002;110(2):245-56.
123. Sharpe NA, Tepper JM. Postnatal development of excitatory synaptic input to the rat neostriatum: an electron microscopic study. Neuroscience. 1998 Jun;84(4):1163-75.
124. Mullah SH, Inaji M, Nariai T, Momose-Sato Y, Sato K, Ohno K. Optical analysis of developmental changes in synaptic potentiation in the neonatal rat corticostriatal projection. Neuroscience. 2012 Jan 10;201:338-48. Epub 2011 Nov 17.
125. Araki KY, Sims JR, Bhide PG. Dopamine receptor mRNA and protein expression in the mouse corpus striatum and cerebral cortex during pre- and postnatal development. Brain Res. 2007 Jul 2;1156:31-45. Epub 2007 Apr 22.
126. Inaji M, Sato K, Momose-Sato Y, Ohno K. Voltage-sensitive dye imaging analysis of functional development of the neonatal rat corticostriatal projection. Neuroimage. 2011 Feb 1;54(3):1831-9. Epub 2010 Oct 13.
127. Pellegrino LJ, Altman J. Effects of differential interference with postnatal cerebellar neurogenesis on motor performance, activity level, and maze learning of rats: a developmental study. J Comp Physiol Psychol. 1979 Feb;93(1):1-33.
128. Diaz-Granados JL, Greene PL, Amsel A. Memory-based learning in preweanling and adult rats after infantile x-irradiation-induced hippocampal granule cell hypoplasia. Behav Neurosci. 1992 Dec;106(6):940-6.
129. Wilson DA, Sullivan RM, Leon M. Single-unit analysis of postnatal olfactory learning: modified olfactory bulb output response patterns to learned attractive odors. J Neurosci. 1987 Oct;7(10):3154-62.
130. Gamse JT, Kuan YS, Macurak M, Brösamle C, Thisse B, Thisse C, Halpern ME. Directional asymmetry of the zebrafish epithalamus guides dorsoventral innervation of the midbrain target. Development. 2005 Nov;132(21):4869-81. Epub 2005 Oct 5.
131. Kuan YS, Yu HH, Moens CB, Halpern ME. Neuropilin asymmetry mediates a left-right difference in habenular connectivity. Development. 2007 Mar;134(5):857-65. Epub 2007 Jan 24.
132. Kuan YS, Gamse JT, Schreiber AM, Halpern ME. Selective asymmetry in a conserved forebrain to midbrain projection. J Exp Zool B Mol Dev Evol. 2007 Sep 15;308(5):669-78.
133. Beretta CA, Dross N, Guiterrez-Triana JA, Ryu S, Carl M. Habenula circuit development: past, present, and future. Front Neurosci. 2012;6:51. Epub 2012 Apr 23.

134. McLean DL, Fetcho JR. Using imaging and genetics in zebrafish to study developing spinal circuits in vivo. Dev Neurobiol. 2008 May;68(6):817-34.
135. McLean DL, Fan J, Higashijima S, Hale ME, Fetcho JR. A topographic map of recruitment in spinal cord. Nature. 2007 Mar 1;446(7131):71-5.

136. Seymour B, O'Doherty JP, Koltzenburg M, Wiech K, Frackowiak R, Friston K, Dolan R. Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. Nat Neurosci. 2005 Sep;8(9):1234-40.
137. Oleson EB, Gentry RN, Chioma VC, Cheer JF. Subsecond dopamine release in the nucleus accumbens predicts conditioned punishment and its successful avoidance. J Neurosci. 2012 Oct 17;32(42):14804-8.
138. Leknes S, Lee M, Berna C, Andersson J, Tracey I. Relief as a reward: hedonic and neural responses to safety from pain. PLoS One. 2011 Apr 7;6(4):e17870.
139. Belova MA, Paton JJ, Morrison SE, Salzman CD. Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. Neuron. 2007 Sep 20;55(6):970-84.
140. Salomons TV, Johnstone T, Backonja MM, Davidson RJ. Perceived controllability modulates the neural response to pain. J Neurosci. 2004 Aug 11;24(32):7199-203.
141. Riemensperger T, Völler T, Stock P, Buchner E, Fiala A. Punishment prediction by dopaminergic neurons in Drosophila. Curr Biol. 2005 Nov 8;15(21):1953-60.