

Benjamin Ehret

Sensory and behavioral correlates of
aversive conditioning in prefrontal
cortex population activity

Diss. ETH No. 28466

SENSORY AND BEHAVIORAL CORRELATES
OF AVERSIVE CONDITIONING IN
PREFRONTAL CORTEX POPULATION
ACTIVITY

A dissertation submitted to attain the degree of
DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by
BENJAMIN EHRET
MSc., University of Zurich and ETH Zurich
born on 06 March 1990
citizen of Germany

accepted on the recommendation of
Prof. Dr. Benjamin Grewe, examiner
Prof. Dr. Fritjof Helmchen, co-examiner
Prof. Dr. Valerio Mante, co-examiner

2022

Abstract

To survive in challenging environments, animals need to learn to perform adaptive behaviors when exposed to sensory information that predicts threats or rewards. The medial prefrontal cortex (mPFC) has been suggested to link sensory and behavioral aspects of such associations during learning and behavior execution. During learning, sensory-driven prefrontal responses emerge as stimuli gain behavioral relevance, and it has been shown that manipulations of mPFC activity can drive and inhibit the execution of learned behaviors. Yet, how prefrontal circuits map relevant sensory signals onto specific behaviors is still unclear. Here we investigate the involvement of mPFC in mediating such mappings using aversive auditory conditioning paradigms with dynamic stimulus-response associations. We trained mice in two complementary active avoidance experiments in which we manipulated the link between stimuli and conditioned behaviors. We achieved this manipulation by either changing the shock-predictive stimulus or changing the action required for avoidance, which required animals to learn a new mapping between sensory input and behavioral output in both cases. During these paradigms, we used miniaturized microscopes for calcium imaging in freely behaving mice to record the activity of mPFC neurons throughout learning. We employed a decoding approach to quantify how tones and fear-related behaviors are represented in mPFC activity, and how these representations change with learning. We found that prefrontal tone responses were tightly coupled to the behavioral relevance of a given stimulus. Moreover, avoidance actions could be predicted from mPFC activity up to three seconds before action onset. However, we did not find a clear link between the representations of tone-evoked activity and avoidance-predictive activity. These findings stand in contrast to related work that suggests joint coding of tone and behavior as a mechanism for linking sensory information to behavior execution. Taken together, our results motivate theoretical research into how links between sensory responses and action initiation signals could be achieved without such joint coding.

Zusammenfassung

Um in komplexen Umgebungen überleben zu können, müssen Tiere lernen ihr Verhalten anzupassen, wenn Reize in ihrer Umgebung Gefahren oder Belohnungen vorhersagen. Eine Vielzahl an Studien suggeriert, dass der mediale präfrontale Kortex (mPFC) eine wichtige Rolle für die Verbindung von sensorischen Informationen und erlernten Verhaltensweisen spielt. Im Verlauf von Lernprozessen, in denen ein ursprünglich neutraler Stimulus für ein Tier an Bedeutung gewinnt, beginnen Neurone im mPFC auf Präsentationen dieses Stimulus zu reagieren. Zudem wurde gezeigt, dass durch Manipulation der Aktivität im mPFC erlernte Verhalten ausgelöst oder gehemmt werden können. Dennoch ist es weiterhin unklar wie präfrontale Netzwerke relevante sensorische Signale und spezifische Verhalten miteinander verbinden. In dieser Arbeit untersuchen wir die Rolle des mPFC in solchen Verbindungen mithilfe von aversiven auditorischen Konditionierungs-Verfahren in denen sich die Assoziation zwischen Stimuli und Verhalten über die Zeit verändert. Wir trainierten Mäuse in zwei komplementären Active Avoidance Experimenten, in denen die Tiere lernen einen elektrischen Schock zu vermeiden indem sie ein definiertes Verhalten als Antwort auf einen Ton Stimulus ausführen. Um das Erlernen von verschiedenen Assoziationen zwischen Tönen und Verhalten zu untersuchen, veränderten wir im Verlauf der zwei Experimente entweder den Ton, der den Schock vorhersagt, oder das Verhalten, das zum Vermeiden des Schocks nötig ist. Während dieser Verhaltensexperimente nahmen wir mit einem Miniatur-Mikroskop die Aktivität von Hunderten von mPFC Neuronen via Calcium Imaging auf. Mithilfe einer Dekodierungs-Analyse quantifizierten wir, wie Töne und Verhalten in der aufgenommenen mPFC Aktivität repräsentiert sind, und wie sich diese Repräsentationen über den Lernprozess verändern. Unsere Resultate zeigen, dass präfrontale Ton-Aktivität eng an die Verhaltens-Relevanz eines Stimulus gebunden ist. Weiterhin konnten wir Vermeidungs-Verhalten bis zu drei Sekunden vor der Initiierung des Verhaltens anhand der aufgenommenen neuronalen Aktivität vorhersagen. Dennoch haben wir keinen klaren Zusammenhang zwischen Ton-basierter Aktivität und Verhaltens-voraussagender Aktivität gefunden. Diese Ergebnisse stehen im Kontrast zu ähnlichen Studien, die suggerieren, dass die Verbindung von sensorischen Informationen und Verhalten über die gemeinsame Kodierung von Ton-basierter und Verhaltens-bezogener Aktivität entsteht. Zusammengenommen motivieren unsere Resultate theoretische Arbeiten, die untersuchen wie die Transformation von sensorischer Aktivität zur Verhaltenskontrolle ohne solche gemeinsame Kodierung erreicht werden kann.

Acknowledgements

I am deeply thankful to the many people that have - directly or indirectly - contributed to this work over the years. I feel very fortunate to have been surrounded by so many talented, hard-working, caring and supportive people. First of all, I would like to thank Christian, with whom I've shared most of the journey through the PhD. It was a true pleasure to work with such an incredibly competent, reliable and fun colleague and friend through so many phases and different projects. I want to thank Roman for his countless expert contributions to this work, for always being there and all the coffee breaks, and for working hard to create a great social environment. Thank you to Maria for being an amazing colleague in all dimensions, for all the great science, but even more for the emotional support and her presence through the many good and bad times we shared. I want to thank Liz Ann for being both a critical scientist and a supportive and motivating mentor at the same time. Finally I'd like to thank my thesis committee for guiding me through the process. Most importantly, thank you to Benni for teaching me the ways of experimental neuroscience, for his endless positivity and for his trust and commitment to my project. Thank you also to Valerio and Fritjof for valuable feedback and guidance over the years.

I am deeply grateful to my family and friends who have laid out the basis that allowed me to pursue this scientific endeavor. First and foremost I want to thank my parents for their constant and selfless support. Thanks to my mom for her positivity, sense of curiosity and fascination for nature, and to my dad for teaching me to question everything and to think for myself. Thanks to Beffi, Daniel and Julia for always being there and to Sarah and Arianna for sharing this intense time with me and supporting me.

Finally, I want to thank the many people I have met during my time at INI. This place was my home for many years and I have encountered a great variety of interesting people, many of which became good friends.

Contents

1	Introduction	1
1.1	Learning in neural systems	1
1.1.1	Model systems	1
1.1.2	Conditioning paradigms	2
1.1.3	Experimental strategies	3
1.1.4	Recording techniques	4
1.1.5	Population level analysis of neural activity	5
1.2	Neural circuits for associative learning and adaptive behaviors	6
1.2.1	The fear circuit	6
1.2.2	Prefrontal circuits in sensation and behavior	6
1.2.3	Neural circuits involved in active avoidance	7
1.3	Challenges and open questions addressed in this thesis	8
2	Methods	10
2.1	Calcium imaging with miniaturized fluorescence microscopes	10
2.1.1	Surgical procedures	10
2.1.2	Imaging experiments	12
2.1.3	Histological validation of recording location	12
2.2	Behavioral setup for active avoidance experiments	14
2.2.1	Basic components	14
2.2.2	Control software and dynamic gating of stimuli	14
2.2.3	Data acquisition and synchronization	15
2.2.4	Automated platforms as flexible safe zones	16
2.2.5	Analysis of behavior recordings	16
2.3	Signal extraction from calcium imaging movies	18
2.3.1	Preprocessing	19
2.3.2	Signal extraction using PCA/ICA	20
2.3.3	Session alignment	22
2.3.4	Joint processing of multiple sessions	23
2.3.5	Annotation	24
2.4	Analysis of behavioral and neural recordings	26
2.4.1	Behavior quantification	26
2.4.2	Analysis of single cell responses	26
2.4.3	SVM decoding	27
3	Fear Conditioning	31
3.1	Behavioral paradigm	31
3.2	Behavior quantification	32
3.3	Single cell correlates of tones and freezing	33
3.4	Population-level analysis of tone-evoked activity	35
3.5	Single-trial tone representations and their relation to freezing behavior	37

4	Two-dimensional Active Avoidance	39
4.1	Behavioral paradigm	39
4.2	Behavior quantification	40
4.3	Imaging data set	43
4.4	Tone-evoked activity	44
4.5	Avoidance-related activity	47
4.6	Relating tone and avoidance coding	51
5	Two-tone Active Avoidance	53
5.1	Behavioral paradigm	53
5.2	Behavior quantification	55
5.3	Single cell tone responses and decoding of tone-related information . .	57
5.4	Tone identity decoding	58
5.5	Avoidance action prediction	59
6	Discussion	61
6.1	Evolution of prefrontal tone responses over fear conditioning and extinction	61
6.2	Neural correlates of tones and actions in active avoidance	63
6.3	Lack of action-specificity in the 2DAA paradigm	64
6.4	Representational similarity of avoidance-inducing tones	65
6.5	Limitations and outlook	66
6.6	Conclusion	67
	Bibliography	68
7	Appendix	76

Chapter 1

Introduction

1.1 Learning in neural systems

For most animals, the ability to learn how to interact with their environment is key for survival. A sensory stimulus can for example indicate the presence of food or the proximity of a predator and should thus trigger approach or avoidance responses. While such stimulus-response mappings can be innate, a learning mechanism substantially increases flexibility and enables animals to adapt to changing environments. The complexity of the associations and behaviors that an animal can learn varies across species, but many aspects of learning and the underlying neural substrates are conserved across evolution [1, 2]. Historically, the field of behaviorism studied how animals adapt their behavior based on the association of stimuli or actions to positive or negative consequences [3, 4]. Experimental work on animal behavior led to important advances in the theoretical understanding of behavioral learning processes [5], but as the brain was treated as a black box, the neural basis of these processes remained unclear. However, technological advances have led to the availability of brain-internal signals and have been continuously increasing the precision at which neural systems can be investigated, ranging from the level of brain areas to individual neurons, synapses and molecules. How the processes on these levels can be experimentally investigated with respect to their relevance to learning and behavior, is highly dependent on the used model systems and technologies. For example, to address the neural basis of human behavior, one is usually restricted to noninvasive technologies due to ethical considerations. In general, there is a trade-off between the complexity of learning that can be investigated in a model system and the precision of the technologies that can be used to analyze it. In the following sections, I will introduce aversive conditioning in mice as the model system used in this thesis, as well as experimental strategies to study this system. I then cover background information and related work on the relevant neural circuits and methodologies to set the stage for the questions addressed in this thesis. As the work contained in this thesis was conducted by multiple people, I will use 'we' instead of 'I' for the remainder of the thesis. Contributions of different collaborators are clearly indicated in the corresponding section at the end of the document.

1.1.1 Model systems

While understanding human intelligence is arguably most relevant to society, it is also the hardest area to study due to its complexity and ethical restrictions on experimental strategies. Therefore animal models are of tremendous importance to the field of neuroscience, as they allow studying neural systems in greater detail. Many mechanisms are well conserved across evolution and various fundamental

insights come from a diverse set of model organisms, such as the analysis of the action potential in squids [6], the synaptic basis of memory in aplysia [7] or the processing principles of the visual system in cats [8]. In modern day neuroscience, many questions are studied using rodents and especially mice as they fulfill various desiderata. First, as mammals, mice are evolutionary relatively close to humans and their nervous system is structured in a similar way, although key differences exist [9]. Second, the availability of genetic tools enable powerful experimental strategies as further introduced in sections 1.1.3 and 1.1.4. And third, mice are efficient to breed and maintain in a standardized and reproducible manner. Within the field of systems neuroscience, mice and rats are widely used to study diverse areas such as sensory [10], motor [11], navigational [12] and reward systems [13]. In this thesis we are mainly concerned with the circuits for associative learning in the mouse. In the following section, we will describe how these circuits are studied in behaving animals through conditioning tasks.

1.1.2 Conditioning paradigms

In conditioning paradigms animals are trained to form associations between stimuli (Pavlovian conditioning) or to modify their behavior (instrumental conditioning) through rewards or punishments. Here, we briefly cover the concepts related to these two types of conditioning and introduce fear conditioning and active avoidance as examples of aversive paradigms for classical and instrumental conditioning respectively.

Pavlovian conditioning Classical conditioning was first studied by Pavlov in his work with dogs in the late 19th century [3]. In a typical setting, animals learn to associate a neutral conditioned stimulus (CS) with an appetitive or aversive unconditioned stimulus (US). Before conditioning, animals respond to the US with an unconditioned response (UR, e.g. salivating in response to food presentation), while the CS does not induce a behavioral response. By repeatedly pairing CS and US presentations, animals learn an association between the two stimuli, and develop a conditioned response (CR) to presentations of the CS alone, despite its initially neutral nature. Repeated presentations of the CS without pairing to the US lead to the extinction of the conditioned response.

Fear conditioning Auditory fear conditioning, is a widely used version of Pavlovian conditioning [1]. In this paradigm, the CS takes the form of a tone, while the US is an electrical foot shock. After conditioning, CS presentations lead to a behavioral response called freezing, which is characterized by an immobile, crouching posture [14]. Due to its simplicity and robustness, this paradigm has been highly influential in the study of the neural circuits involved in associative learning, as further described below.

Instrumental conditioning In instrumental conditioning paradigms, animals are trained to increase or decrease the frequency of a given behavior through reinforcement or punishment, respectively. Both reinforcement and punishment can be positive (presentation of an appetitive/aversive stimulus), or negative (removal of an appetitive/aversive stimulus). In contrast to classical conditioning, the presentation

of these stimuli is dependent on the animal's behavior, such that animals learn to adjust their behavior to achieve the desired outcome. A classic example is the delivery of food rewards in response to a lever press in a Skinner box [4].

Active avoidance In the context of aversive conditioning, a commonly used instrumental conditioning paradigm is auditory signalled two-way active avoidance (referred to as active avoidance for the remainder of the thesis). In this paradigm, mice are placed in a two-compartment box and are presented with tone stimuli that are followed by aversive foot-shocks, similar to the auditory fear conditioning paradigm. However, in contrast to fear conditioning, mice can avoid receiving the shock by shuttling from one compartment to the other. Through this contingency of stimulus (tone), response (shuttling) and outcome (shock or no-shock), animals learn to perform the instrumental shuttle behavior in response to the tone. While the paradigm was already developed in the 1930s, it was largely abandoned for many years, due to conceptual ambiguities and difficulties to theoretically explain a variety of experimental results [15]. The main conceptual issue is the lack of a clear reinforcement signal, as a successful trial is characterized by the omission of the US. In an early attempt to resolve this paradox, two-factor theory proposed that avoidance learning consists of two phases, a Pavlovian one followed by an instrumental one [16]. The theory postulated that animals acquire a negative association to the CS in the Pavlovian phase, such that the termination of the CS serves as a (negative) reinforcement signal in the instrumental phase. However, there were many experimental results that are inconsistent with this idea [17]. Later, Bolles questioned the instrumental nature of the shuttling action [18]. He viewed shuttling as a species-specific defense reaction (SSDR), similar to conditioned freezing in fear conditioning, that is learned through Pavlovian processes and not via reinforcement. However also this interpretation is challenged by many findings, for example through experiments on active avoidance based on actions that are not SSDRs (e.g. lever pressing) [17]. Further theories consider learned expectations [19, 20] and safety signals [21] for the generation of the reinforcement signal that would give rise to the instrumental learning of the avoidance action. Overall no theory can unify the wide range of experimental results, and there is a consensus that avoidance learning is based on a variety of different learning processes [22]. However, over the last decade, a better understanding of Pavlovian conditioning and novel experimental approaches have sparked new research on active avoidance. A finding that speaks for the instrumental nature of the avoidance response is that avoidance learning relies on neural circuits known to be crucial for the learning of goal-directed appetitive behaviors [22], as further described below.

1.1.3 Experimental strategies

To investigate the relation between neural activity and behavior, two key complementary strategies exist. We can either record neural activity and relate it to aspects of an animal's behavior, or manipulate the neural system and observe the effects that manipulations have on behavior. Manipulation studies allow to investigate causal links between neural activity and behavior, while recording strategies only enable correlative analysis. Historically, lesion studies have been used extensively to determine the brain areas that are necessary for learning and executing different types of behaviors. Another manipulation approach is based on the pharmacological inactivation of neural activity (e.g. through Muscimol), which temporarily disrupts the

function of targeted regions and allows assessing their relevance to a given behavioral process. Over the last two decades, the development of optogenetics [23] and other approaches based on genetic manipulations (e.g. DREADDs [24]), has dramatically increased the specificity of manipulations of neural activity. Such approaches are not used in this thesis, but they have been crucial in analyzing the neural circuits involved in associative learning as further introduced in section 1.2.

Recordings of neural activity during behavior allow investigating the relation of brain-internal signals to sensory inputs and behavioral outputs and the transformations in between. In particular, continuous recordings over multiple days enable the analysis of changes in cellular properties and their relation to learning. In the following sections we introduce different methods for recording neural activity and consider the challenges associated with the analysis and interpretation of the recorded data.

1.1.4 Recording techniques

To record the activity of individual neurons in awake animals, one can either use electrophysiological or optical recording techniques and both approaches have their advantages and disadvantages. Common **electrophysiological approaches** use extracellular electrodes for single-unit or multi-unit recordings and directly measure the electrical activity caused by action potentials evoked in the surrounding cells at high temporal resolution [25]. While this approach has led to countless neuroscientific advances, it has key limitations [26]. First, it is difficult to record from the same cells over multiple days. Second, the cell type and connectivity of a recorded neuron is not easily identifiable. And third, the number of neurons that can be recorded from simultaneously is substantially lower than in imaging approaches. Although the third point saw recent improvements with the development of novel high-density electrodes [27], the first two still limit the applications of electrode recordings.

Over the last two decades, **calcium imaging** has emerged as a powerful alternative to electrode recordings with complementary strengths [26]. Using fluorescent calcium indicators as an indirect measure of neural action potentials, large populations of up to thousands of neurons can be recorded simultaneously using fluorescence or excitation microscopy [28]. Due to their stable position in the microscope field of view, individual neurons can be identified over multiple recording sessions enabling long-term recordings. Additionally the use of calcium indicators enables diverse strategies to tag specific cell subsets, e.g. based on their genetic identity or projection specificity via genetic and viral strategies. However, one key limitation of calcium imaging approaches is the low temporal resolution originating from the slow kinetics of the used calcium indicators (see [29] for a review). Imaging neuronal fluorescence signals requires optical access to the investigated tissue. Cranial windows can be used to make the surface of the brain visually accessible through a cover glass [30]. Two-photon microscopy allows imaging up to several hundred micrometers deep from the brain surface [31]. However, many deep brain regions remain inaccessible using this approach. This issue can be circumvented by the use of gradient index (GRIN) lenses, that can be implanted directly above the region of interest, such that deep brain signals are relayed outside of the animal's skull [32].

One restriction that results from the use of regular microscopes is that animals have to be head-fixed below the microscope for the recording of neural signals. This imposes severe limitations on the study of neural signals relating to behavioral paradigms, which often require animals to move freely. One approach to study motion-based behaviors using two-photon microscopes is the use of virtual reality systems [33, 34, 35]. In such a setup, animals are head-fixed, but can navigate through a virtual environment presented on a screen by moving on an air-supported spherical treadmill. An alternative approach that allows for actual motion is based on miniaturized head-mounted microscopes [36, 37]. A widely used system that is also employed in this thesis is combining such miniature microscopes with GRIN lenses implanted above the area under investigation [37, 38, 39].

1.1.5 Population level analysis of neural activity

The possibility to simultaneously record the activity of hundreds of neurons results in the challenge of analyzing this high-dimensional neural activity [40]. In early sensory areas, the activity of individual neurons can often be understood in terms of simple stimulus features [8]. However, activity in higher-order brain areas gets increasingly difficult to interpret. For example, a cell might be responsive to multiple different variables relevant to a given task [41].

Encoding approaches, often based on generalized linear models (GLMs), attempt to reconstruct the activity of individual cells through linear combinations of different predictive variables, such as stimuli, behaviors or the activity of other neurons [42, 43, 44]. Populations of neurons can then further be analyzed in terms of the parameters of these encoding models.

Conversely to the encoding approach, **decoding approaches** make use of pattern recognition methods such as support vector machines (SVMs) for assessing what kind of information is represented in the recorded neural population activity [41, 45]. In particular, differences in the performance of decoders trained at different time points can reveal dynamic processes that change the information content contained in neural recordings. Cross-temporal analysis of the different decoders can further give an indication of how information is encoded at different time points [46]. In a similar way, one can compare decoders trained with data from different recording sessions to assess changes in neural coding over longer time scales, such as changes associated with learning [39].

A more direct way to investigate dynamic processes unfolding on the population level is the use of **dimensionality reduction techniques** [47, 48, 49]. This approach facilitates the interpretation of high-dimensional neural activity, since single cell activity can often only be understood when analyzed in relation to the activity of other cells in the recorded population [50, 51]. In many cases, the low-dimensional representation resulting from dimensionality reduction constitutes a good summary of population-level dynamics and can be related to the computations involved in a given task [52]. To further facilitate interpretation, information about task-variables can be directly integrated into the dimensionality reduction process [51, 53]. Such techniques can be used to obtain state-space dimensions that contain information specifically related to aspects of a task that are of particular interest.

Finally, population dynamics can be analyzed by fitting **latent factor models** in the form of dynamical systems that capture the processes underlying the variability in neural population recordings. Over the last years such fitting procedures have been

introduced for linear [54], switching linear [55] as well as nonlinear dynamical systems [56].

1.2 Neural circuits for associative learning and adaptive behaviors

1.2.1 The fear circuit

Studies on the neural basis of fear conditioning identified the amygdala as the key region for learning CS-US associations [1]. The amygdala consists of multiple sub-nuclei such as the basolateral amygdala (BLA) and the central amygdala (CeA). The connectivity and physiology of different cell types and microcircuits have been extensively studied in relation to sensory stimuli and behavioral outputs [57]. In a simplified picture, CS-US associations are learned in the BLA [39, 58], while CeA circuits mediate the expression of freezing behavior via projections to the periaqueductal grey (PAG) [2, 59, 60]. However, the acquisition of fear memories and the expression of fear-related behaviors generally depend on a distributed network of brain areas [61, 62]. In particular, the medial prefrontal cortex (mPFC) is thought to provide top-down control signals to the BLA [63, 64, 65]. Within the mPFC, the prelimbic (PL) and infralimbic (IL) subregions, are associated with the expression of fear, and the regulation of extinction, respectively. Manipulation studies have demonstrated opposing effects on freezing when stimulating PL (increased freezing) and IL (decreased freezing) [66]. It has further been shown that PL activity is necessary for the expression of learned fear using pharmacological [67] and optogenetic approaches [68]. Using electrode recordings, Burgos-Robles et al. [69] showed that fear conditioning induces tone responses in PL, and that the time course of these conditioned responses correlates with freezing behavior. As such correlations had not been observed in the lateral amygdala, which drives PL responses, the authors suggested that mPFC transforms transient sensory-evoked signals into sustained behavioral outputs which are fed back into the basal amygdala. mPFC and BLA are bidirectionally connected [61, 70] and coordinated activity between the two areas has been shown to regulate fear behavior [71] and to mediate between conflicting motivational drives [72]. Courtin et al. [73] investigated the mechanistic basis of prefrontal involvement in fear expression and found that local interneurons organize mPFC output to the BLA to drive behavior execution. During fear expression, the coordination of neural activity in mPFC and BLA is based on a 4Hz rhythm [74] and the manipulation of this rhythm in mPFC can bidirectionally modulate fear expression [75]. Taken together, these results show that mPFC receives information about threat related stimuli and is involved in producing freezing behavior. However, it remains unclear if and how threat related activity is transformed within mPFC to drive behavior, especially in the context of the observed 4Hz rhythm.

1.2.2 Prefrontal circuits in sensation and behavior

Outside of its role in fear conditioning, the rodent PFC has been studied in the context of various other learning processes linking sensory inputs and behavioral outputs (see [76, 77] for reviews). One common motif found in different types of conditioning is that prefrontal cells only display substantial responses to sensory stimuli, once they have gained a behavioral relevance, e.g. through predicting a reward or punishment [69, 77, 78]. Le Merre et al. [79] show that the strength of sensory responses correlates with task performance and that mPFC inactivation significantly

decreases performance. These results indicate a role for sensory-driven mPFC signals in behavior execution, but how a transformation from sensory signals to behavioral outputs is implemented is still unclear. mPFC's role in behavior execution has been studied through optogenetic manipulation of projection-specific subpopulations that relay prefrontal signals to subcortical areas involved in behavior execution such as the nucleus accumbens (NAc), the paraventricular thalamus (PVT), the BLA or the PAG. Using this approach, mPFC has been shown to be involved in the execution of social [80, 81], defensive [68, 82, 83] and reward-based behaviors [77, 78]. Yet it remains challenging to relate activity in these subpopulations to sensory driven activity. Otis et al. [78] show that mPFC neurons projecting to the NAc and PVT are biased to display positive and negative stimulus-evoked responses respectively. Further, manipulation of these projection-specific subpopulations has opposing influences on behavior execution. These findings provide a first link between mPFC stimulus representations and their relation to driving behaviors. Yet it is difficult to assess a transformation from stimulus-evoked activity to behavioral outputs due to task-inherent overlaps of stimulus presentation and behavior execution.

Another line of research investigates mPFC's role in the learning of rules and categories. While such functions are well established for primate PFC [84], they have only recently been described for mPFC in rodents [85, 86]. In the context of links between stimuli and behavioral output, such rule and category representations are interesting as they are usually investigated in terms of changing mappings from stimuli to responses. For example, in the work of Reinert et al. [86], mice were trained to classify visual stimuli into two categories based on different stimulus features by licking on one of two reward spouts. Category selectivity was then defined based on differences in a neuron's activity in response to stimuli within and between categories. The presence of such category selectivity again shows that mPFC neurons represent sensory inputs in a manner that is geared towards their behavioral relevance. However it remains unclear if and how these contextualized sensory representations are used to drive specific behaviors. Work on cognitive flexibility shows that different actions [87] and strategies [88, 89] are correlated to different activity patterns in mPFC neurons. Yet, how such behavior related activity relates to sensory input is still unclear. Overall these results are consistent with a model in which mPFC attributes learned meaning to sensory information and guides the selection of appropriate behaviors in an efficient and flexible manner, but the mechanisms of such transformations remain to be understood.

1.2.3 Neural circuits involved in active avoidance

The learning and execution of active avoidance behavior is mediated by a network of brain areas associated with both classical and instrumental conditioning [90]. In particular NAc, BLA and mPFC have been implicated in orchestrating the balance between passive and active fear-induced behaviors [91]. Dopamine signalling in the NAc predicts avoidance [92] and inactivation studies have shown that the NAc is required for the execution of avoidance actions [90, 93]. Lesions or inactivation of the BLA resulted in similar effects on avoidance expression [90, 93, 94], establishing the necessity of this area. During avoidance learning, animals have to suppress freezing in order to perform avoidance actions. Consistent with this notion of suppression, lesions of the CeA, which is crucial for the expression of freezing, do not affect or even promote avoidance [94]. A recent study further added to the understanding of

this circuit by showing that divergent PVT projections to CeA and NAc control the balance between the opposing behaviors of freezing and avoidance [95].

As introduced above, the mPFC projects to BLA, NAc and PVT. In the context of active avoidance, mPFC activity has been shown to be elevated during sustained avoidance [96] and necessary for the expression of avoidance actions [97, 98]. In a recent study, Jercog et al. [99] showed that mPFC tone responses are driven by inputs from the BLA and that mPFC activity predicts avoidance responses. These results indicate that mPFC might transform threat-related information to drive avoidance actions. This view is consistent with a recent optogenetic study that studied the effect of stimulating and silencing mPFC projections to NAc and BLA [100]. The study found that mPFC could bidirectionally modulate avoidance behavior via these two projections. These results suggest a model in which mPFC uses sensory information to orchestrate active and passive defensive behaviors. However, as for fear conditioning and other paradigms, how such a transformation from sensory signal to action initiation signal is implemented is still unclear.

1.3 Challenges and open questions addressed in this thesis

As described above, it is well established that mPFC is involved in the execution of various conditioned behaviors and that behaviorally relevant stimuli evoke responses in mPFC neurons. However, it is still poorly understood how stimulus-driven mPFC activity specifically relates to the execution of behavior. Many studies that investigate the link between sensory-driven activity and behavior combine neural recordings with manipulation approaches to assess the relevance of some form of neural activity to behavior. However manipulations of neural activity are limited in terms of their specificity. While stimulus-evoked activity can be precisely identified in the recordings, it remains challenging to specifically manipulate activity patterns related to stimulus presentations. This lack of precision restricts the specificity of statements regarding causal relations between activity and behavior. For example, when inhibiting or stimulating an entire brain area, it is unclear whether the resulting effects on behavior are caused by the manipulation of specific activity patterns related to sensory stimuli or rather by disrupting a network that is generally important for the regulation of behavior.

The specificity of manipulation techniques can be increased using different strategies such as the targeting of genetically or anatomically defined cell types, or the use of temporally precise manipulations. For example, as previously mentioned, Otis et al. [78] manipulate projection-specific subsets of prefrontal neurons to demonstrate opposing effects of behavior. To link these opposing effects to stimulus-related activity, the authors perform projection-specific imaging and show that there is a bias in how cells in the different subpopulations respond to a conditioned tone stimulus. While these results are important, the links between stimulus-related activity and behavior that can be demonstrated with this approach are limited to effects that align with the used manipulation and recording strategies.

In this thesis we use a complementary approach to address these limitations. Instead of manipulating neural activity while observing changes in stimulus-induced behavior, we manipulate the stimulus-response mapping that an animal learns through conditioning and observe the concurrent changes in neural activity. This approach allows us to investigate the nature of stimulus-response mappings in mPFC by observing changes in neural activity upon learned changes in the association between

sensory inputs and behavioral outputs. We train mice on different conditioning paradigms with dynamic stimulus-response mappings and record prefrontal neural activity using miniature fluorescence microscopes for calcium imaging over multiple days. This technology is well suited, as it enables recording large numbers of cells in freely moving animals. Crucially, the ability to track the activity of individual cells over the full course of an experiment allows us to evaluate changes in coding over the time scales associated with learning.

We start with a classic fear conditioning experiment to investigate how stimulus-related and behavior-related activity interact in mPFC and change over the course of learning and extinction. We then design two active avoidance paradigms in which we change the mapping between stimuli and instrumental behaviors over time. In the first experiment, we change the behavior induced by a single tone, whereas in the second experiment, we change the tone which induces a single behavior. By quantifying the effect that these manipulations have on mPFC activity, we can make statements about the nature of the link between stimuli and behaviors. To identify and quantify neural signals related to stimuli and behaviors we employ population-level decoding approaches. By analyzing and comparing neural coding at different time points, we gain insights into the temporal evolution of mPFC representations over learning. Finally, we compare and interpret the results of our different experiments to obtain a clearer picture of how mPFC is involved in linking stimuli and different types of behaviors.

Chapter 2

Methods

2.1 Calcium imaging with miniaturized fluorescence microscopes

In this section we will introduce the methodology of calcium imaging with head-mounted miniature fluorescence microscopes, and describe the experimental procedures we used to acquire the data sets that are considered in this thesis. All active avoidance data sets were acquired as described in this chapter, while the fear conditioning data set was acquired by Benjamin Grewe in Stanford, using closely related procedures as described in [39]. Below, we will detail how mice were prepared for imaging experiments, how these imaging experiments were conducted, and how we validated the source of the recorded signals by post-mortem analysis of the considered brain tissue.

2.1.1 Surgical procedures

Calcium imaging experiments require the use of calcium indicators that emit light in response to neural activity, as well as optical access to the tissue under consideration. Here we briefly describe our imaging approach before detailing all procedures below. There are various different calcium indicators and approaches for introducing them into the neurons of interest. Which approach to use depends on multiple factors such as the considered cell type and brain region, the duration of the imaging experiments or the required temporal resolution (for a review see [29]). Here, we chose to work with GCaMP6m [101], which we introduce into prefrontal excitatory neurons using a viral vector that is locally injected through a craniotomy. Recording the fluorescence signals emitted by the used calcium indicators requires optical access to the investigated brain region. Since the prelimbic subregion of the mPFC, which we investigate in this thesis, is located approximately 2 mm below the pia, the light signals of interest are obscured by the tissue of more superficial brain regions as well as the mouse's skull. To obtain access to these deep-brain signals, we implant a GRIN lens into the mouse's brain, which relays the optical signals from PL to the outside of the animal's skull, where they can be imaged by a microscope. Finally, as we are working with head-mounted miniaturized fluorescence microscopes, we attach a base plate to the mouse's skull, which allows mounting the microscope for imaging sessions and detaching it again afterwards.

All procedures were approved by the Veterinary office of the Kanton Zurich (Cantonal Veterinary Office Zurich; active avoidance experiments) or in accordance with the regulations of the Stanford Administrative Panel on Laboratory Animal Care (fear conditioning experiment).

Subjects. Experiments were performed on male C57BL6/J mice (Jackson Labs US; fear conditioning experiment) and C57Bl6/Crl1 mice (Charles River Germany; active avoidance experiments), between 4 to 7 months old at the time of the behavioral experiment. Animals were housed in individually ventilated cages (IVC) in a 12h light/dark cycle room (lights on from 7:00 am to 7:00 pm), and were provided food and water ad libitum. After import from the breeders, mice were given a 2 weeks period to acclimatize to the new housing condition prior to the first surgery. Mice used in active avoidance paradigms were kept in groups of 2 to 5 animals. For fear-conditioning experiments, mice were individually housed at least 14 days prior to the start of experiment.

Anesthesia. For all procedures including anesthesia mice received preemptive buprenorphine (Bupaq; Streuli, 0.1 mg/Kg) 20-30 minutes prior to anesthesia. Anesthesia was induced with a Ketamin-Xylazin cocktail (Ketanarcon; Streuli, 90 mg/Kg / Xylazin; Streuli, 8 mg/Kg), and mice were subsequently mounted onto a stereotactic frame (Kopf Instruments). During the procedure mice received 95% medical oxygen (Pangas, Conoxia) through a face mask and body temperature was kept steady at 37 degree Celsius using a temperature controller and a heating pad.

Viral injections. At the time of the virus injection surgery, mice were 8 - 12 weeks old. To label excitatory neurons in the prelimbic cortex, we intracranially injected 500nl of an adeno-associated virus driving the expression of GCaMP6m via the CamKII promoter (AAV2/5-CamKIIa-GCaMP6m; Schnitzer-lab) into the prelimbic cortex (1.8 mm anterior, 0.4 mm medial, 2.1 mm ventral relative to bregma). The virus was injected either via a micropump (UMP3UltraMicroPump; WPI) or via a borosilicate glass pipette with a 50 μ m diameter tip connected to a picospritzer (Parker), using short pressure pulses at a speed of approx 100 nl/min. After injection, the needle was left in place for 5 minutes to avoid back-spill and finally the skin was closed using suture.

Microendoscope implantation. 7 to 14 days after the viral injection, we implanted a small stainless steel guide tube (1.2 mm diameter; Ziggy's tubes and wires) with a custom glass coverslip (0.125 mm thick BK7 glass, Electron Microscopy Science) glued to the bottom end, above the injection site (c.f. [38, 37]). In brief, we first made a round craniotomy (1.3 mm diameter) centered above the prelimbic cortex (1.8 mm anterior, 0.4 mm medial relative to bregma). To avoid increased intracranial pressure when inserting the implant, we aspirated tissue down to a depth of 1.9 mm from the skull surface. Next, we lowered the guide tube to the bottom of the incision (2.2 mm relative to skull surface) and glued the guide tube to the mouse skull using UV curable glue (4305 LC: Loctite). We then applied dental cement (metabond; Parkell or Scotchbond ESPE; 3M) over the complete cranium and around the guide tube. Finally we attached a metal bar and applied dental acrylic cement (Paladur) to build up and stabilize the implant. Animals received 6-10 weeks recovery time before testing viral expression levels.

Selection of animals for imaging. To assess the quality of the preparation, we head-fixed mice on a running wheel and inserted GRIN lens into the guide tube (GT-IFRL-100-101027-50-NC; GRIN-Tech). We then placed the miniature microscope (nVista HD 2.0, Inscopix) over the GRIN lens and visually assessed the quality of the recorded signals. We only used animals if viral expression lead to considerably high

fluorescence levels that allowed us to clearly resolve the activity of distinct neurons.

Base plate mounting. Approximately one week before starting behavioral experiments we prepared mice for imaging with the miniature microscope. First, we anaesthetised animals and mounted them onto a stereotactic frame (Kopf Instruments). Next, we glued a GRIN lens into the guide tube to allow permanent access to neural signals. We then precisely positioned the microscope (base plate attached) above the GRIN lens to find the optimal focus plane. Finally, we attached the microscope base plate to the skull using blue light curable glue (Pentron; Flow-it) and detached it from the microscope.

2.1.2 Imaging experiments

Habituation to the microscope. Prior to behavioral experiments, we habituated all mice to the microscope mounting procedure and the weight of carrying the microscope for at least three consecutive days. During the mounting procedure, animals were briefly head-fixed on a custom made mounting station with a running disc using their metal head bar. We then attached the microscope to the base plate and let the animal freely explore an open arena within the experimental room for 10 minutes. Additionally, subjects were handled for five out of seven days by the experimenter.

Imaging parameters. During the habituation procedure, we determined the optimal recording parameters for the nVista miniscope. First, we adjusted the microscope focus to optimize image quality. The focus setting of the nVista miniscope is based on a screwing mechanism that rotates the camera sensor to increase or decrease its distance to the specimen. Through this rotation, the focus setting affects the orientation of the recorded images with respect to the mouse's orientation. We only worked with 90 degree rotations and noted the sensor's orientation to be able to relate the orientation of the recorded images to the orientation of the mouse brain in our analyses. Next, we selected a 1000×1000 pixel field of view from the available 1440×1080 pixel space according to the local quality of the preparation. We then set the LED intensity to the minimal value that produced sufficiently bright images. The required intensity depends on the strength of GCaMP6m expression and high LED power leads to undesirable photobleaching, which has to be accounted for in the analysis of the resulting movies (cf. section 2.3.1). Over all subjects, we used an LED intensity between 10% and 25% ($0.2-0.5 \text{ mW/mm}^2$), while the imaging sensor gain level was always set to its maximum value of 4.

Data acquisition. Before each recording session we carefully chose the field of view to match the one determined during the habituation procedure to facilitate between session alignment. Start, end and synchronization of the recording was controlled by the behavioral setup as described in section 2.2. We acquired frames of 1000×1000 pixels with 16 bit precision at a frame rate of 20 Hz. All recorded data was directly streamed to the hard disc of the recording computer.

2.1.3 Histological validation of recording location

After recordings were completed, mice underwent terminal anesthesia with Pentobarbital (Esconarkon; Streuli, 200 mg/Kg). We transcardially perfused all mice with

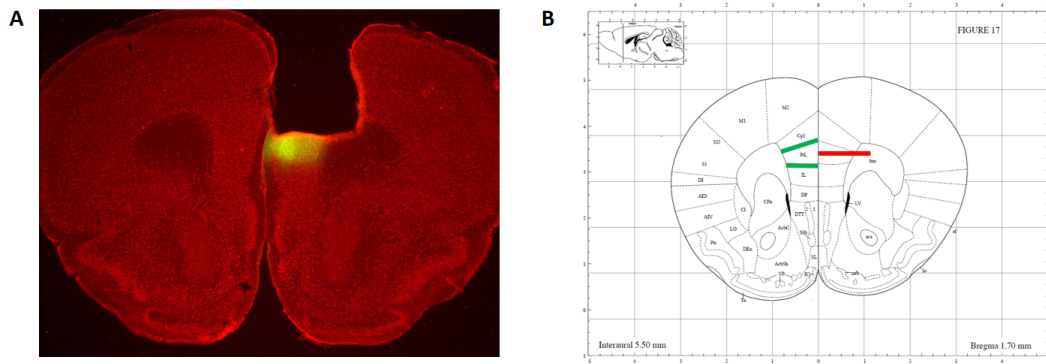


FIGURE 2.1: GRIN lens implant location for an example mouse. **(A)** Coronal brain slice with Nissl stain (red) and GCaMP6m expression (green) below the hole left by the implanted microendoscope. **(B)** Estimated location (red bar) of the implant in PL (borders indicated in green), overlaid on the appropriate coronal section of the Paxinos reference atlas [102].

PBS followed by cooled paraformaldehyde (4% PFA). Next, we removed brains from the cranium and stored them in PFA for 24-28 hours. We then used a Vibratome (VT2000s, Leica) to cut 50 μm thick coronal brain slices and stored them in PBS. To highlight cytoarchitectural structures in the prepared slices, we performed a Nissl stain to label cell bodies with Alexa Fluor. Next, we mounted all slices on microscope slides and acquired large field-of-view images with a standard fluorescence microscope (Z16, Leica) using two channels for GCaMP6m and Alexa. Finally, we used a reference brain atlas [102] to estimate the position of the endoscope in the recorded images as shown for an example animal in Fig. 2.1. Most implants were situated in PL, but due to slight variations we use the term mPFC throughout the thesis.

2.2 Behavioral setup for active avoidance experiments

In this section we describe the experimental setup we built for conducting different types of active avoidance experiments. The experimental setup of our fear conditioning experiments is detailed in [39]. We start by describing the basic components and control mechanisms used for a standard active avoidance paradigm and later explain the additional features we added to enable more complex experimental designs (see sections 4.1 and 5.1).

2.2.1 Basic components

In a standard active avoidance paradigm, a mouse is placed in a two-compartment box and learns to shuttle between the two compartments in response to a tone, in order to avoid receiving an aversive foot shock (for a detailed explanation see section 4.1). We used a combination of commercially available hardware and custom build extensions and control software to create a flexible active avoidance setup that we used in various settings. To minimize the influence of external stimuli, we built the setup in a ventilated and illuminated isolation box (Campden Instruments, 4120-6, displayed in Fig. 2.2). We used a shuttle box with two shock grids (Coulbourn, H10-11M-SC), that were individually controlled by two precision shockers (Coulbourn, H13-16). The two compartments of the box were originally separated by a wall with a small opening. We replaced this wall with a small hurdle in order to allow mice to shuttle between the two compartments with a head-mounted microscope, while maintaining the compartmentalization. Tones were supplied at 75 dB by a pair of stereo speakers (Logitech, Z2000) and the inside of the shuttle box was recorded by two behavior cameras with overlapping fields of view (ImagingSource, DMK 23FV024). We controlled all experiments via a Windows 10 PC running control software written in MATLAB (The MathWorks, 2017a). The PC was connected to actuators and sensors via a National Instruments PCI card (PCIe-6323) and a data acquisition device (BNC-2090A). To dynamically control stimuli during an experiment, we used an Arduino microcontroller (Nano 3.0) and a custom designed PCB (printed through Eurocircuits) as described in the following section.

2.2.2 Control software and dynamic gating of stimuli

To automatically control stimulus presentations and data acquisition, we wrote a custom MATLAB software package (code available at https://gitlab.ethz.ch/henningc/fear_conditioning_setup, access upon request). In this package, the structure of an experiment is determined by a design file, which specifies the timing of tone and shock presentations. During an experiment, these predefined signals are sent out via the National Instruments Data Acquisition board (NIDAQ). In order to ensure temporal precision of all signals, outputs are buffered onto the NIDAQ. However, in the active avoidance paradigm, we need to adjust stimulus presentations based on the subject's behavior. In particular, we need to determine in which compartment to supply the shock based on the subjects position at the beginning of a trial, and we need to be able to dynamically suppress tones and shocks when the subject performs the shuttle behavior. To achieve this, we developed an additional hardware component that acts as a gate between the scheduled output signals of the NIDAQ and the speakers and shockers that receive these signals. Using this gating system, we route shocks to the appropriate compartment for each trial, and block tones and shocks whenever the subject shuttles from one compartment to the

the Arduino outputs at 1000 Hz and recordings are precisely aligned to the start and end of an experiment. This high recording rate and precise alignment allowed us to downsample NIDAQ recordings to match the 20 Hz frequency of the movie recordings.

2.2.4 Automated platforms as flexible safe zones

In conventional active avoidance experiments mice are trained to shuttle between the two compartments of the shuttle box. The shock grids in these compartments can be controlled individually, to make sure that animals that escape the shock in one compartment immediately notice the absence of the shock in the other compartment (i.e. there is no delay through a shuttle detection system that shuts off the shock). This hardware implementation limits the possibilities to compartmentalize the shuttle box to the compartments defined by the shock grids. To enable more complex experimental designs based on changing the compartmentalization of the shuttle box, we developed an automated platform system, that allowed us to flexibly define safe zones during active avoidance trials (Fig. 2.3). This system comprises four 3D-printed platforms, that tile the cage and can be lifted up between the bars of the shock grid, such that animals can stand on the platform without touching the shock grid (Fig. 2.3A,C). We build a lifting mechanism located below the shock grid using LEGO technic (Fig. 2.3B). The lifting mechanisms of the four platforms were automatically and individually controlled by four motors that operate pulling cables which lift or lower the platforms (Fig. 2.3D). Motors were located outside of the isolation box and controlled by the Arduino. The exact use of these platforms for the two-dimensional active avoidance paradigm is detailed in section 4.1.

2.2.5 Analysis of behavior recordings

We recorded animal behavior within the shuttle box using two top-view cameras with slightly overlapping fields of view (Fig. 2.4A,B). To analyze these recordings, we first stitched the two videos together. We first cut out the overlapping regions in the two videos and stretched each row of pixels using interpolation to recover the original frame sizes (Fig. 2.4C). To track the location of the mouse in the stitched video, we used DeepLabCut [103]. We used the tool to track the positions of 5 marker points throughout all behavior videos: miniscope top, miniscope base, left ear, right ear and tail base (Fig. 2.4D). For analyses which required the use of a single tracking point, we calculated the mouse centroid as the mean of the positions of all markers excluding the miniscope top. An example motion trace of this centroid is displayed in Fig. 2.4E.

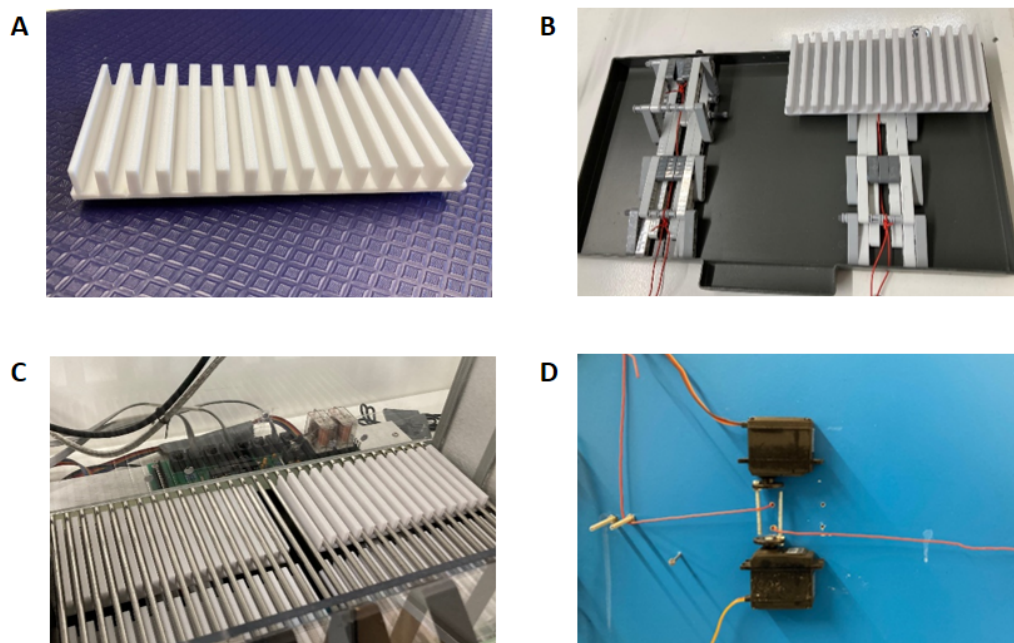


FIGURE 2.3: Photographs displaying automated platform system. **(A)** 3D-printed platform that fits between the bars of the shock grid. **(B)** Platform raising mechanism based on Lego technic. Platforms can be individually raised by pulling on the red cables. In the displayed picture we only mounted one of the four platforms to illustrate the lifting mechanism. The top left stage is in the raised state, all other stages are in the default state. **(C)** Platform system attached below shock grid. Upper right platform is in the raised state, such that an animal can stand on it without touching the shock grid. **(D)** Motors attached to the outside of the isolation cage, which operate the pulling mechanism via the red cabled.

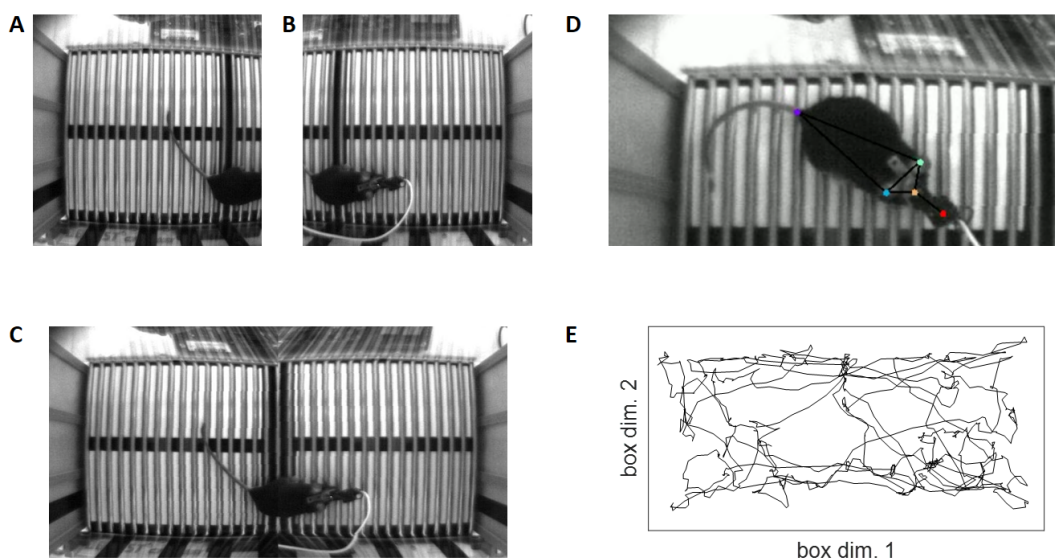


FIGURE 2.4: Analysis of behavior camera recordings. **(A,B)** Simultaneously recorded example frames of the two behavior cameras which have overlapping fields of view. **(C)** Result of the stitching procedure **(D)** DeepLabCut tracking points: miniscope top (red), miniscope base (orange), left ear (green), right ear (blue) and tail base (purple). **(E)** Motion trace of a 5 minute window displaying exploratory behavior.

rease has sparked research on ways to automatically extract neural signals from calcium imagin

2.3 Signal extraction from calcium imaging movies

Recent advances in calcium imaging technologies have drastically increased the amount of data that can be recorded in a typical experiment. This incg movies in order to deal with large data sets (for reviews see [104, 105]). The challenge of signal extraction is to correctly assign light intensity changes in the recorded movies to individual cells and to distinguish a given cell's activity from other sources of light (in the following referred to as noise). The nature of this noise depends on the features of a given experiment, such as the cytoarchitecture of the investigated brain region or the used calcium indicator and imaging technology. For example, the analysis of 2-photon microscopy data [106, 107] differs from the analysis of 1-photon microscopy data [108, 109], as 1-photon measurements include more out-of-focus fluorescence signals due to the bigger integration volume as compared to 2-photon measurements. Existing analysis approaches differ in the way they deal with these sources of noise. While one approach is to remove all undesired signals in a preprocessing step [110], other approaches explicitly model different sources of variability with appropriately constrained models [106, 108]. However, this modelling comes at the expense of additional computational costs.

Here, we chose to use the former approach in order to keep signal extraction computationally efficient to allow the processing of large data sets. This efficiency offers an elegant solution to a challenge occurring in experiments that consist of multiple recording sessions: tracking individual cells over multiple days. Movies that are acquired over different sessions can differ in the captured field of view (through e.g. translations or rotations), and the identified cells need to be matched between recordings. The classical way to achieve this is to identify cells per day, and to then attempt to match as many cells as possible between sessions based on their location (c.f. [39]). Here, we follow a different approach: we first align and concatenate movies from different recording sessions, and then perform signal extraction once for the joint movie. This approach has two key advantages: First, data annotation requirements are reduced from once per session to once per subject, which amounts to a factor of up to 11 in our data sets and second, for every identified cell, we obtain an activity trace over the full duration of an experiment. This is usually not the case when matching cells identified in individual sessions. For example, a given cell might be identified in some sessions, but not in others, e.g. due to low activity levels or variability in the signal extraction algorithm or the annotation process. In contrast, this issue does not occur with the joint analysis of multiple aligned sessions.

In the following sections, we will describe the steps from raw imaging data to validated neural activity of cells that we can follow over all recording sessions. We start with the preprocessing process, where we detail how different sources of noise are addressed. We then introduce the PCA/ICA algorithm, explain how to align recordings from different sessions and show how to jointly process them in a memory-efficient way. Finally, we cover data annotation procedures to validate the automatically extracted neurons. An implementation of this processing pipeline is available at <https://gitlab.ethz.ch/behret/movieanalysis> (access upon request).

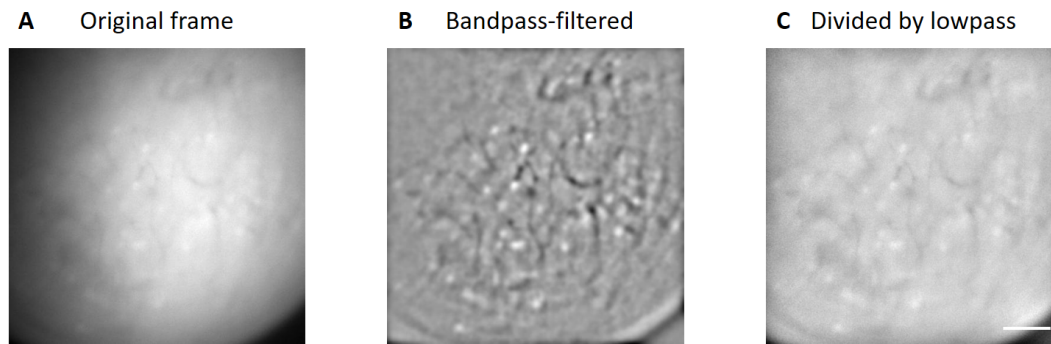


FIGURE 2.5: Illustration of filtering steps used in preprocessing. **(A)** Raw output of the nVista microscope. **(B)** Band-pass filtered version of (A). Structural features such as blood vessels are highlighted to facilitate pairwise registration of frames for motion correction (cutoff frequencies: 30-80 pixels). **(C)** Frame from (A) divided by its low-pass filtered version. Low-pass filtering isolates signals with a large spatial extent, which are then removed from the original frame through division (cutoff frequency: 7 pixels). Scale bar: 100 μm .

2.3.1 Preprocessing

In all experiments we started the analysis with the raw output files of the nVista 2.0 HD microscope, which contained 1000×1000 pixel images recorded at 20 Hz. In a first step, we concatenated all recording files belonging to one session and down-sampled each frame by a factor of two, resulting in movie stacks of size $500 \text{ pixel} \times 500 \text{ pixel} \times n$, where n is the number of frames recorded in a given session. Subsequently we applied the following preprocessing steps to prepare the recorded data for signal extraction.

Motion correction and temporal downsampling To account for small translations of the field of view originating from motion of the head-mounted microscope, we used an established image registration method (Turboreg [111]). To align all frames, we randomly selected a reference frame and performed pairwise registration with all remaining frames. To facilitate registration, we applied a spatial band-pass filter to every frame, which enhances constant structural features such as blood vessels (see Figure 2.5). We used cutoff frequencies of 30 and 80 Hz for the band-pass filter for all experiments. After obtaining registration coordinates using the band-pass filtered frames, we used these coordinates to align the frames of the original movie. Next, we cropped all pixels on the edges of the field of view, for which information was lost in frames that had to be corrected. Finally, we downsampled the corrected movie in time by a factor of 4 for computational efficiency, resulting in a 5 Hz frame rate.

Debleaching Photobleaching leads to a decrease in signal intensity, in particular for long imaging sessions and high LED power values. This signal decrease needs to be accounted for in order to avoid the contamination of neural activity signals. One way to address spatially uniform changes in signal intensity is to normalize each frame using filtering techniques. However, we observed that bleaching differently affects individual pixels in the recorded movies, especially in relation to blood vessel patterns (see appendix for details and illustration in Fig. 7.1). To address this issue, we built a low-rank model of the slowly changing pixel patterns associated with bleaching and subtracted it from the movie. To build this bleaching model,

we first represented the movie in matrix form by collapsing the two pixel dimensions. We then temporally smoothed the movie with a window size of 400 seconds to highlight the slow changes associated with bleaching and discard variability with high temporal frequency (e.g. neural activity). Next, we used principle component analysis (PCA) to determine the pixel dimensions which are most affected by slow changes in pixel intensity (see Fig. 7.2). We then built a rank-2 bleaching model by reconstructing the movie in the space spanned by the first two principle components. Finally, we subtract the bleaching model from the original movie, to discard the slowly changing variance captured by the rank-2 reconstruction.

Filtering Out of focus fluorescence causes changes in luminosity that have a larger spatial extent than the signals of individual neurons. These luminosity fluctuations can be isolated in individual frames using a spatial low-pass filter. Thus, to remove wide-field luminosity fluctuations, we divided each individual frame by its low-pass-filtered version (see figure 2.5). We used a cutoff frequency of 7 Hz for all experiments.

DF/F In a final preprocessing step before cell extraction, we re-expressed all frames in units of relative changes in fluorescence, given by $\Delta F(t)/\bar{F} = (F(t) - \bar{F})/\bar{F}$, where \bar{F} is the mean frame obtained by averaging over the entire movie.

2.3.2 Signal extraction using PCA/ICA

Here we will give a brief introduction to the PCA/ICA algorithm from Mukamel et al. [110] to explain how we used it for the joint analysis of long term recordings from multiple sessions in the following sections. The goal of the algorithm is to assign the variance in the preprocessed calcium imaging movies to independent sources (i.e. neurons). These sources are represented as pairs of a spatial map and a temporal trace, where the map indicates how individual pixels contribute to a given source, and the trace indicates when the source was active (see Fig. 2.6). Independent component analysis (ICA) is a source separation algorithm that can provide such a decomposition, but it is too computationally intensive to be applied to calcium imaging movies directly. However, principal component analysis (PCA) can be used as a preparatory step to compress and denoise the movie data, allowing further processing with ICA.

As PCA and ICA operate on matrices, we first represent the movie in matrix form by collapsing the two pixel dimensions and obtain

$$M \in \mathbb{R}^{n \times t} \tag{2.1}$$

where n is the number of pixels and t is the number of frames. We perform PCA to reduce dimensionality in the time dimension, i.e. we have n observations given by the pixels and t variables given by the frames. Principle components can be computed via the singular value decomposition (SVD) of M , or via the Eigendecomposition of the covariance matrix of M . Here we choose the latter version, as it is more efficient when the number of observations exceeds the number of variables, which is the case for our data (usually on the order of 250000 pixels vs. 10000 frames).

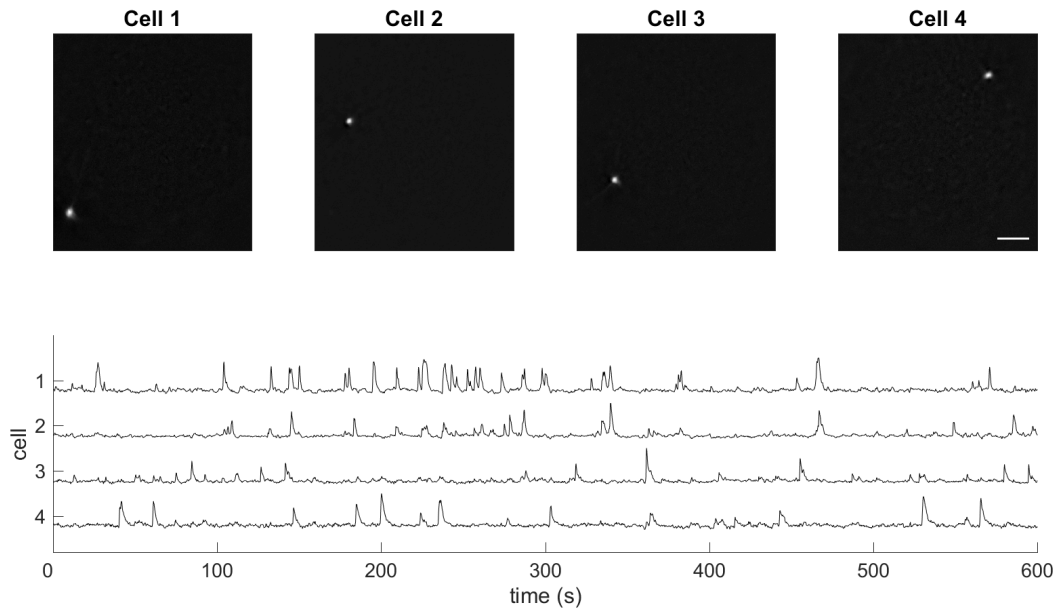


FIGURE 2.6: Illustration of PCA/ICA outputs. **Top:** Pixel maps of 4 example cells indicating cell location in the field of view. Scale bar: $100 \mu\text{m}$. **Bottom:** 10 minute sample of the activity traces of the 4 example cells.

Principle component analysis To perform PCA, we first mean-subtract each frame of M to obtain the matrix M' and calculate the covariance matrix C_M

$$C_M = M'^T M' \quad (2.2)$$

where we omit the normalization by the number of samples, which is handled below. We then calculate the Eigendecomposition of C_M

$$C_M = Q \Lambda Q^{-1} \quad (2.3)$$

where Q is a $t \times t$ matrix that contains the Eigenvectors of C_M and Λ is a diagonal matrix containing the corresponding Eigenvalues. To reduce dimensionality, we can consider the matrix

$$T \in \mathbb{R}^{t \times k} \quad (2.4)$$

which consists of the first $k \leq t$ columns of Q , representing the Eigenvectors that span the k -dimensional space in which M has maximal variance. We then obtain the first k principle components of M , by projecting M' onto the space spanned by T .

$$S = M' T \Lambda^{-\frac{1}{2}} \quad (2.5)$$

where S is a $n \times k$ matrix, whose columns contain the ordered PCs, and the normalization with the singular values $\Lambda^{\frac{1}{2}}$ ensures that all PCs have a uniform variance of $\frac{1}{n-1}$.

Independent component analysis The matrices T and S already have the desired output format displayed in figure 2.6, as the columns of S correspond to pixel maps that describe the spatial structure of the k components, whereas the columns of T correspond to the intensity of these components over the t frames. However, the

components computed by PCA still mix signals from different neurons. To demix these signals, ICA computes a linear transformation matrix

$$W \in \mathbb{R}^{k \times l} \quad (2.6)$$

where $l \leq k$ is the predefined number of independent components. As the k principle components are defined through their spatial and temporal features reflected by S and T , Mukamel et al. [110] use a spatiotemporal mixing matrix

$$X = \begin{bmatrix} (1 - \mu)S \\ \mu T \end{bmatrix} \quad (2.7)$$

with dimensions $(n + t) \times k$ as input to ICA, where the parameter μ determines the relative importance of spatial and temporal information. In usual applications of ICA, whitening of the input matrix is an important preprocessing step that ensures that all dimensions are treated equally by ensuring the input dimensions are uncorrelated and have a variance of 1. Importantly, the former is already the case for S and T . Furthermore, for both matrices all dimensions have uniform variance and differences between the two variances are absorbed by μ when combining spatial and temporal information. To calculate W from X , we use the fastICA algorithm [112]. We then obtain the final spatial components S' by demixing the spatial components S with W .

$$S' = SW \in \mathbb{R}^{n \times l} \quad (2.8)$$

Finally, we obtain the demixed temporal traces T' through the projection

$$T' = M^T S' \in \mathbb{R}^{t \times l} \quad (2.9)$$

For further processing, we restructure the spatial components in S' such that S' becomes a tensor with dimensions $x \times y \times l$, where x and y are the number of pixels of the two dimensions of the recorded frames.

Parameter settings In accordance with Mukamel et al. [110] and based on exploration with our data, we used a value of $\mu = 0.1$ for all experiments. Based on the quality of the individual calcium imaging movies, we set the number of independent components l between 400 and 600, roughly corresponding to 1.5 times the expected number of neurons. We then set the number of principle components k to $1.2 \times l$.

2.3.3 Session alignment

Since the position of the miniature microscope is slightly different in every recording session, the position of a given cell in the recorded field of view will not be the same for different recordings. However, the relative position of all recorded cells stays constant over multiple imaging sessions. Thus, to align two recordings, we can use the positional information of the extracted cells contained in S' . To obtain a compact representation of the cells' locations, we create a cell map of size $x \times y$ by performing a maximum projection over the third dimension of the tensor S' . Figure 2.7 shows the cell maps of two example sessions and visualizes the transformation between them.

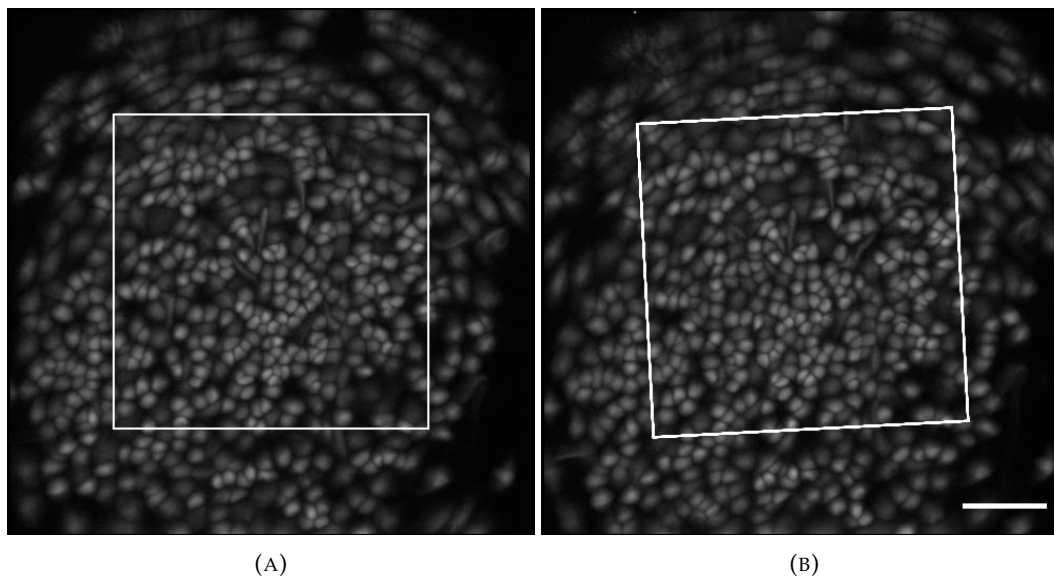


FIGURE 2.7: Illustration of cell maps used for session alignment. **(A)** Cell map of reference session. The white frame illustrates the transformation applied to this cell map (none for reference session). **(B)** Cell map of example session that needed to be rotated to align with the reference session. The necessary rotation is illustrated by the white frame. Scale bar: $100 \mu\text{m}$

To align all recording sessions of a subject, we collect the corresponding cell maps and align all of them to a reference session. We usually choose session $\lfloor (\frac{s}{2}) \rfloor$ as the reference session, where s denotes the total number of sessions. If the alignment based on this reference session is not satisfactory, we explore other reference sessions until we find a good alignment. To align pairs of cell maps, we use MATLAB's *imregister* function. Note that the feasibility of such an alignment highly depends on the quality of the recorded data. In a final step we visually validate all alignments and exclude sessions for which proper alignment was not possible from further analysis.

Using the session alignments, we can concatenate the recordings of all sessions for joint processing. However, the movies of different recording sessions might have different signal to noise ratios. To account for these differences, we calculate the overall standard deviation of all pixels for every session and then scale every session to match the minimal standard deviation for a given subject. The resulting concatenated movie then contains DFOF values with a stable mean and standard deviation over all sessions.

2.3.4 Joint processing of multiple sessions

Using the concatenated movie of all sessions of a subject, we can rerun signal extraction with PCA/ICA to obtain a common cell map over the full experiment. This has the benefit, that we track cells over all sessions by design. This is not the case when performing signal extraction per session on then matching cells between sessions. If, for example, a cell is rarely active in some sessions but highly active in others, it might only be identified in some sessions and can not be analyzed over the full experiment. With our approach on the other hand, we follow the activity levels of all cells throughout all recorded frames. However, this process creates challenges due to the time and space complexity of the computations involved in PCA/ICA.

The first problematic step is the calculation of the covariance matrix C_M in equation 2.2. For individual sessions, we can simply load a movie into working memory and calculate C_M . When concatenating the movies of multiple sessions, already loading all data becomes problematic. The size of a movie is given by

$$s_m = \frac{x \times y \times t \times 4}{1024^3} \text{GB} \quad (2.10)$$

where x and y are the sizes of the two pixel dimensions, t is the total number of frames and where every pixel is represented as a 4 byte floating point number (MATLAB's data type *single*). In our longest experiment (11×12000 frames in total), s_m equals approximately 123 GB. Additionally, the size of the covariance matrix is given by

$$s_c = \frac{t^2 \times 4}{1024^3} \text{GB} \quad (2.11)$$

and thus increases quadratically with the number of frames, totalling approximately 64 GB for our longest experiment. As our computing hardware is constrained to 128 GB RAM, we circumvent loading the full concatenated movie by splitting the computation of the covariance matrix into batches of 12000 frames. Instead of computing C directly, we compute all possible

$$C_{i,j} = M_i'^T M_j' \quad (2.12)$$

where $i, j \in \{1, \dots, b\}$ are the batch indices and b is the total number of batches of the mean-subtracted movie M' . C can then be simply assembled by concatenating all $C_{i,j}$ matrices. In our example, this procedure reduces the memory requirements for movie data to approximately 22 GB, as only two batches have to be kept in working memory at a time.

The second problematic step is the Eigendecomposition of C in equation 2.3. This computation requires increased precision (from 4-byte single to 8-byte double), increasing the original memory requirements s_c by a factor of two. Additionally, the computation of the Eigendecomposition of s_c requires the same amount of RAM as holding s_c in memory. Thus s_c is limited to one fourth of the available working memory, i.e. 32 GB in our case. As the size of the full covariance matrix exceeds this limit, we have to perform the calculation of the spatial components S' with a subset of all available frames. However, we can recover the activity traces over the full duration of the concatenated movie, by projecting the full movie onto these spatial components (see equation 2.9). To combine data from all sessions, we use a subset of 6000 consecutive frames from each session for the calculation of S' , yielding a manageable covariance matrix.

Postprocessing A known issue with PCA/ICA is that individual cells are occasionally split into multiple components if the number of independent components is not set perfectly. To make sure we do not include split cells in our analyses, we remove cells that have highly correlated activity (Pearson correlation > 0.7 and are spatially close (centroid distance < 20 pixels)).

2.3.5 Annotation

As a final step, we manually annotate the components resulting from joint extraction for every subject. To judge whether a component accurately reflects the activity of

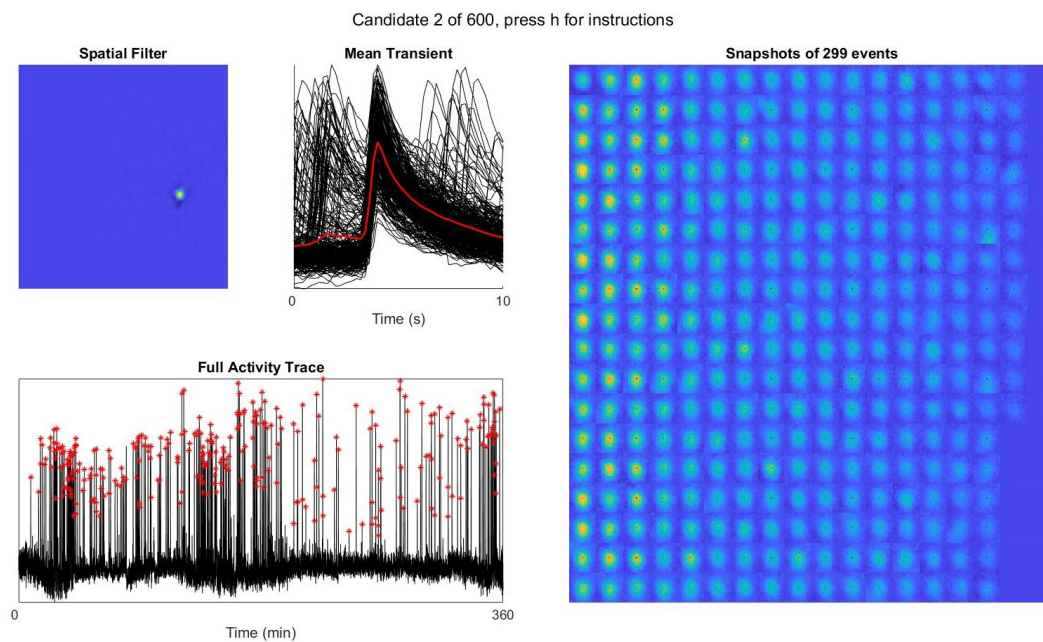


FIGURE 2.8: Screenshot of the annotation software we use for validating all cells. The panels display the key characteristics that we use to judge if a candidate cell is valid (see text). **Top left:** Pixel map. **Bottom left:** Activity trace with detected events marked by red crosses. **Top middle:** 10s cutouts around detected events (black) and averaging over all aligned events (red). **Right:** Event snapshots given by cutouts around the cell location at the time of detected events.

a cell over the full experiment, we make use of the following information (see Fig. 2.8):

- **Component shape:** The shape of the extracted spatial component needs to be consistent with the typical size of a neuron in the recorded movie.
- **Stability of overall trace:** The extracted activity traces needs to have a stable base line over all sessions and should ideally display clear events.
- **Temporal evolution of identified events:** We detect calcium events as peaks in the activity trace and construct a mean event by aligning windows around these peaks and averaging over events. The temporal evolution of the resulting mean event should be consistent with the kinetics of the used calcium indicator (fast rise, exponential decay).
- **Consistency of event-snapshots:** We collect cutouts around the cell centroid for the frames corresponding to detected events. These cutouts visualize the cause of peaks in the activity trace and should consistently resemble the spatial component representing the cell's shape.

We provide examples for violations of the listed characteristics in the appendix. All potential neurons that are not validated through this procedure are excluded from further analysis.

2.4 Analysis of behavioral and neural recordings

2.4.1 Behavior quantification

We analyzed speed and freezing in the same way for all three experiments.

Analysis of motion. We used the markers of the DeepLabCut tracking algorithm to calculate the speed of mice during an experiment. To obtain one general speed value, we first calculated a centroid based on all markers except the one for the top of the miniscope. Next we used the x, y position of the mouse centroid to calculate the instantaneous speed of the animal for every frame as

$$s_t = \sqrt{\text{grad}(x_t)^2 + \text{grad}(y_t)^2} \quad (2.13)$$

where $\text{grad}(\cdot)$ is the numerical gradient and t is the frame index. For some analyses we evaluated the motion of individual markers and the speed of each marker was calculated in the same way as for the centroid. We measured speed in pixels per second.

Freezing classification. We classified time periods as freezing based on sustained immobility. We first smoothed the centroid using a Gaussian window with a standard deviation of 10 frames and then we applied a threshold of 5 pixels/s. Time periods were classified as freezing if the speed was below the threshold value for at least 2 seconds.

Freezing scores. To quantify tone-induced freezing in the fear conditioning paradigm for individual animals, we calculated a tone specificity score and a tone discrimination score. The tone specificity score (SS) was defined as

$$SS = \frac{\text{freezing}_{CS+}}{\text{freezing}_{CS+} + \text{freezing}_{\text{outside-CS+}}} \quad (2.14)$$

such that a value of 1 indicates that a mouse only displayed freezing during the CS+, and a value of 0 indicates that there was no freezing during the CS+.

Similarly, we defined the tone discrimination score (DS) to compare freezing induced by the two tones as

$$DS = \frac{\text{freezing}_{CS+}}{\text{freezing}_{CS+} + \text{freezing}_{CS-}}. \quad (2.15)$$

Avoidance action onsets. For every avoidance trial, we defined the onset of the avoidance action using the speed of the animal. We calculate the speed increase for every time step in the 2s window preceding the recorded shuttle time and define the action start as the time step before the maximal increase in speed.

2.4.2 Analysis of single cell responses

Tone-responsive cells. In all three experiments we defined cells as tone-responsive using the following procedure: We first averaged tone-presentations over trials obtaining the trial averaged tone responses $a(t)$ (grouping trials over sessions, tasks or

overall). To account for the variability of trial length in the active avoidance experiments, we only used trials that were sufficiently long for the t individual time steps. Next, we defined a pre-tone period with the same length as the tone presentations as baseline window (25s for fear conditioning, 10s for avoidance experiments). We then calculated z-scores as

$$z_{\text{tone}}(t) = \frac{a(t) - \mu_{BL}}{\sigma_{BL}}. \quad (2.16)$$

where μ_{BL} and σ_{BL} are the mean and standard deviation over the trial averaged baseline window. We then calculated the average z-score over the full tone presentations and classified neurons as tone-responsive if the absolute value of their average z-score exceeded a threshold value of 2. To calculate tone-onset z-scores and define tone-onset-responsive cells, we used the same procedure, but only averaged tone z-scores over the first 2 seconds of the tone presentation.

Freezing score. In the fear conditioning experiment we calculated a freezing score for each cell. We concatenated data from the 3 post-conditioning days and calculated the freezing score as the Pearson correlation coefficient between the binary freezing trace and a given cell's activity trace.

Speed score. To compare freezing-related activity to speed-related activity in the fear conditioning experiment, we calculated a speed score for each cell. To avoid confounding speed scores with freezing related information, we only used non-freezing time points to assess how a cell's activity related to speed. Since we observed that the relationship between neural activity and speed was not linear, we used a log transform of speed values and then calculated Pearson correlation coefficients between neural activity traces and log-transformed speed traces.

Avoidance score. To isolate avoidance related activity from tone-related activity, we defined the avoidance score in terms of a difference between avoidance and error trials. First, we aligned avoidance trials according to the detected action start (see above). Next, we used the average length of avoidance trials (from tone start to action start) as the alignment point for error trials, such that the trials in both groups had the same average length. For every time step in the 3s window before the alignment points, we then calculated the time-dependent discrimination index as

$$d'(t) = \frac{\mu_a(t) - \mu_e(t)}{\frac{\sigma_a(t) + \sigma_e(t)}{2}} \quad (2.17)$$

where $\mu_a(t)$ and $\mu_e(t)$ are the time-dependent averages over the aligned avoid and error trials, and $\sigma_a(t)$ and $\sigma_e(t)$ are the corresponding standard deviations. Finally, we calculated the overall avoidance score per cell, as the average d' over the 3s window preceding the alignment point.

2.4.3 SVM decoding

To quantify if and how information regarding tones and actions was represented in mPFC activity at different time points, we used a decoding approach based on support vector machines (SVMs) [99]. We trained SVM decoders to discriminate neural activity vectors between two different classes (e.g. tone vs. baseline) and used the

decoding accuracies as a measure of how well information was represented. Before giving details on the different decoding settings, we first explain basic procedures that are common between all settings.

Basic setting. We used SVMs with linear kernels and trained them to classify neural activity vectors of 200ms time steps. We constructed balanced data sets containing time steps from two classes, e.g. 50 time steps during the tone and 50 time steps outside of the tone (baseline). To estimate test accuracies, we used 5-fold stratified cross-validation. If not stated otherwise, we pooled neurons from different subjects to increase decoding performance and to allow for compact analysis of all cells at once. The use of balanced classes and pooling over subjects puts constraints on the number of data points used in the construction of the classification data set. In particular, the maximum number of data points per classes is limited by the minimum number of data points over classes and subjects. In other words, the used number of data points needs to be small enough, such that we have enough examples available for all classes and subjects. To equalize the number of data points between subjects and classes, we subsample data points in cases where more data is available. To deal with the variability induced by subsampling and to estimate the variability of the accuracy measure resulting from decoding, we repeat all decoding analyses using different random samples for each iteration (bootstrapping). For each of the decoding settings below, we specify the number of sample points as well as the number of bootstrap iterations used.

Tone vs. baseline decoding. To analyze tone-related information on individual days, we trained decoders to distinguish time points during tone presentations from time points from the inter-trial interval (baseline). We first built a decoding data set, which is given by a matrix of size $n \times t$, where n is the number of cells (features) used for the decoding, and t is the number of time steps (observations) that should be classified as either tone or baseline. The number of time points in the tone class is limited by the number of tone presentations. Additionally, one issue with the decoding of time series data, is that temporal correlations between time points can introduce confounding information. To avoid this effect, we only used one time step per trial. Thus, t was determined by the number of tone presentations on a given day for each of the experiments (e.g. 50 for the 2DAA experiment), multiplied by 2 (same amount of tone and baseline time steps). For each bootstrap iteration, we randomly sampled one time step from the tone presentations of each trial, and randomly sampled an equal number of time points from the inter-trial interval. We estimated within day accuracy using 5-fold stratified cross-validation. When performing across day testing, we trained models for individual days using all available data, since no cross-validation is needed when testing on data from another day. For all settings, we performed 20 bootstrap iterations.

Trial-wise tone vs. baseline decoding. In the fear conditioning experiment, we tested the presence of tone-related information on individual trials. Since we were interested in the link between tone information and freezing behavior, and freezing behavior differs between different animals, we performed this analysis on individual subjects, rather than using pooled data over subjects. To assess tone information on individual trials we used the following procedure: we pooled all 36 CS+ trials from the 3 post-conditioning sessions and trained 36 different models, always training on 35 trials and testing on the remaining trial, to obtain trial-specific accuracy

values. We randomly sampled 10 time points per trial for computational efficiency and performed 50 bootstrap iterations.

Time-resolved tone vs. baseline decoding. To investigate the time-dependence of tone-related information, we trained decoders for individual time steps. The procedures were the same as for regular tone vs. baseline decoding, but instead of training one decoder on data from randomly sampled time points, we trained individual decoders for every time step. In active avoidance, the number of available trials is different for every time step, as trials can be terminated by avoidance actions. Additionally, avoidance actions provide a confounding factor that can be used to discriminate between tone and baseline settings. To address these issues, we limited the time resolved tone vs. baseline decoding analysis to the first 4 seconds of tone presentations, and excluded trials with an action start within the first 4 seconds. We used 20 trials in the 2DAA and 2TAA per-task models, and 13 trials in the 2TAA per session models. In all per-task models, we performed 20 bootstrap iterations, in per session models we performed 50 bootstrap iterations due to the increased variability resulting from the smaller number of training examples.

Tone ID decoding. In the 2TAA experiment, we asked how well we could distinguish the two different tones at different time points in the same way as for the tone vs. baseline comparison, by replacing the baseline class with data from the second tone. We performed this analysis per session (25 samples, 50 bootstrap iterations), per time step for each task (20 samples, 50 bootstrap iterations) and per time step for individual days (9 samples, 50 bootstrap iterations).

Avoid vs. error decoding. To evaluate if mPFC activity contained predictive information on upcoming avoidance actions, we trained time-resolved decoders to distinguish avoidance and error trials for time steps preceding avoidance action start. As we were interested in differences between the two tasks in both active avoidance experiments, we performed this analysis by pooling data within task sessions, excluding the first learning sessions for both tasks. We aligned avoidance trials using the detected action start time points. For error trials such an alignment condition does not exist. To reduce any bias that might be introduced by avoidance and error trials having different duration, we matched the distribution of tone lengths over trials between error and avoidance trials. This was achieved by sampling alignment points for error trials from the distribution of avoidance alignment points. This sampling was different for every bootstrap run. Using these alignment strategies for the two trial groups, we trained decoders for all time steps from 3s before to 1s after the alignment points. We excluded trials for which avoidance actions started before 3s after tone start. For both the 2DAA and the 2TAA experiments, the number of time steps for the two classes was 10 (minimum number of error trials for one subject) and we performed 100 bootstrap iterations. Next to the described analysis based on the action start alignment, we performed the same procedures to data aligned at tone start (1s before to 3s after tone start) as a control.

To evaluate whether the observed accuracy values differed from values expected by chance we performed a permutation test for the decoding results obtain on every individual time point [113]. We first created a null distribution of chance-level accuracy values by repeating the following procedure for 200 permutations: we randomly permuted the labels of the avoidance and error classes and trained SVMs on the shuffled data for 10 bootstrap iterations and calculated the mean accuracy

over bootstrap runs. We then compared the observed mean accuracy value from the real data to the distribution of mean accuracy values from the 200 permutations and computed p-values as the fraction of permutations whose mean accuracy exceeded the observed value.

Since differences in the neural correlates of motion between avoid and error trials might contribute to decoding performance, we asked how well avoidance actions could be predicted from motion information alone. To do so, we repeated the same analysis as above, but instead of using neural activity as predictive features, we used the speed values computed for the 5 DeepLabCut tracking points.

As we were particularly interested in differences of action-related activity between the two tasks, we evaluated the performance of task-specific decoders across tasks to compare across-task and within-task performance. For across task testing, we trained decoders using all data from one task (no cross-validation), and tested them on all data points from the other task. To further analyze potential task-specificity of action predictive information, we trained a set of decoders (referred to as mix decoders) using an equal amount of data from both tasks. In order to keep results comparable, we used the same number of data points for the mix decoders and the task-specific decoders. In particular, we used half of the data from tasks 1 and 2 for training mix decoders, and used the other half for evaluating their performance on unseen data.

Chapter 3

Fear Conditioning

To investigate mPFCs involvement in linking sensory information and conditioned behavioral responses, we recorded the activity of mPFC excitatory neurons over a six day fear conditioning paradigm. We start this chapter by presenting the behavioral paradigm and quantifying the recorded behavioral data. In the subsequent sections, we first cover basic response properties of single neurons and later analyze how information is represented at the population level throughout the course of the experiment. Finally we analyze how behavioral responses on individual trials are related to sensory-evoked activity.

3.1 Behavioral paradigm

Fear-conditioning experiments were performed in the same manner as described in a previous study from our lab [39]. The paradigm consisted of six imaging sessions over six consecutive days, comprising two habituation sessions, one fear conditioning session and three fear testing sessions (Fig. 3.1A). In each session, mice were presented with multiple repetitions of CS+ and CS- tone pip sequences (25 200 ms pips played at 1 Hz). The tones had a frequency of either 4 kHz presented at 85 dB or 10 kHz presented at 80 dB, and were randomly assigned as CS+ and CS- for each subject in a counterbalanced manner. In habituation and fear conditioning sessions, mice were presented with five CS+ and five CS- tones in an interleaved order starting with CS-. Fear testing sessions started with two CS- presentations, continued with twelve CS+ presentations, and ended with two further CS- presentations. In all sessions the inter-trial intervals had a pseudo-random duration between 20 and 180 seconds. To minimize the level of contextual fear, the fear conditioning session took place in a different chamber (chamber B) as compared to habituation and fear testing

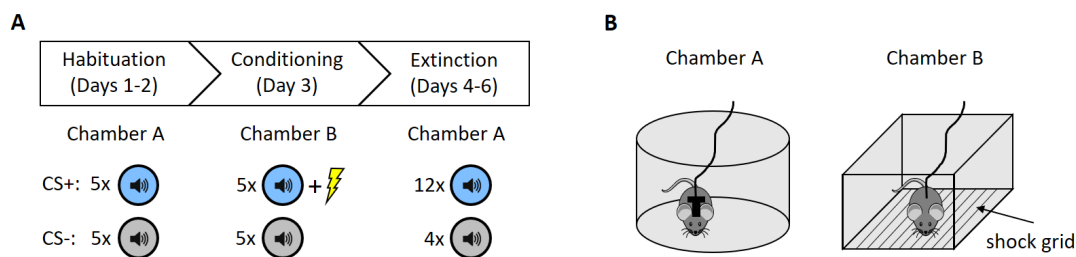


FIGURE 3.1: Fear conditioning paradigm. **(A)** Schedule of the 6-day fear conditioning paradigm. Session-types take place in different boxes and have different tone presentation schedules. **(B)** Schematic of the two behavioral boxes used to minimize contextual fear. Box B was used in the conditioning session and included a shock grid for US presentations.

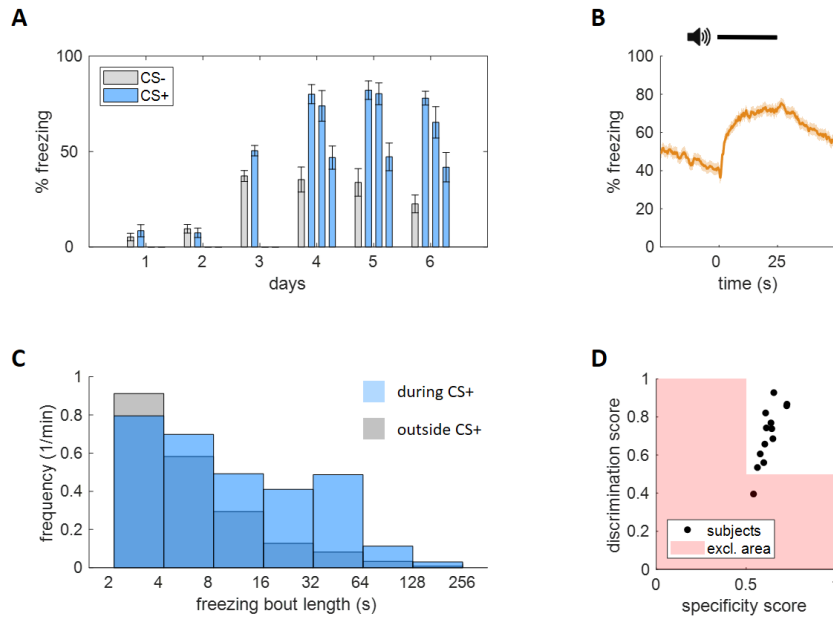


FIGURE 3.2: Fear conditioning behavior. **(A)** Freezing levels during tone presentations over days (Mean \pm SEM, $n=13$ mice). Bars represent groups of 5 (days 1-3) or 4 (days 4-6) tone presentations. **(B)** Freezing probability per time step around CS+ presentations (Mean \pm SEM, $n=13$ mice). **(C)** Distribution of freezing bout lengths over all subjects for bouts starting during, and bouts starting outside of CS+ presentations. **(D)** Freezing specificity and discrimination score per subject. Subject with either score smaller than 0.5 were excluded from further analyses.

sessions (chamber A, Fig. 3.1B). The two chambers had different shape, texture and lighting, and were cleaned with either 1% acetic acid (chamber A) or 70% ethanol (chamber B) to produce different odors. During the conditioning session on day 3, mice received a US foot-shock (2s, 0.6mA) 800 ms after the termination of the last tone pip for each of the five CS+ presentations. Over all sessions, mouse behavior was recorded with a top-view camera.

3.2 Behavior quantification

Overall we recorded data from 14 mice. We had to exclude data from one mouse due to the low quality of its calcium imaging data, which made it impossible to align recordings from different sessions. Another animal was excluded as it did not learn the CS-US association sufficiently well as further explained below. To assess learning-induced changes in animal behavior, we quantified freezing levels on all days of the fear conditioning paradigm (see section 2.4.1). Before the conditioning session on day 3, mice showed very limited levels of freezing during presentations of the CS+ and CS- tones (Fig. 3.2A). In the fear testing sessions on days 4 to 6, animals showed strongly elevated freezing levels during CS+ presentations. Freezing during the CS- was also increased with respect to pre-conditioning sessions, but did not reach the levels of CS+ induced freezing. In all three testing sessions, we observed pronounced within-session extinction. While the first four CS+ presentations lead to an average freezing level of 80.4%, the average dropped to 45.3% for the last four CS+ presentations. Across-session extinction was less pronounced (averages of 66.9%, 69.9% and 61.7% for the 3 post-conditioning days).

To investigate whether freezing was tone-induced, or rather a reflection of a general fearful state, we quantified the freezing probability per time step around CS+ presentations (Fig. 3.2B). While freezing levels were high in general, freezing probability jumped from 45.1% in the 25s before the tone to 66.2% during CS+ presentations, indicating that freezing behavior was driven by auditory input. Freezing probability remained high through the tone presentation and only slowly decayed after tone end. We further compared the length of freezing bouts starting during and outside of CS+ presentations. Freezing bouts that started outside of tone presentations were predominantly short, and long-lasting freezing was rare (Fig. 3.2C). In contrast, freezing bouts that started during CS+ presentations had a markedly increased probability of being of long duration, further suggesting that tone presentations induced elevated freezing levels in post-conditioning sessions.

We next analyzed the individual variability of freezing behavior over all subjects. For all mice, we quantified the CS+ specificity of freezing, both with respect to general freezing, and with respect to CS- induced freezing. We calculated a specificity score, which compared freezing during CS+ presentation versus outside CS+ presentations (see section 2.4.1). A score greater than 0.5 indicated that mice showed higher freezing levels during than outside of tone presentations. As intended by the conditioning paradigm, all mice had a score greater than 0.5 (Fig. 3.2D). Yet none of the subjects came close to the maximum score of 1 (max. score of 0.73), which can be explained by high baseline freezing levels and pronounced post-tone freezing (Fig. 3.2D). Analogously to the specificity score, we calculated a discrimination score, which compared CS+ and CS- induced freezing. All but one animal showed higher CS+ than CS- freezing levels. The animal that had a discrimination score smaller than 0.5 was also the animal with the lowest specificity score, and we excluded this animal from further analysis, as it had not learned the intended stimulus-response association.

3.3 Single cell correlates of tones and freezing

Over all 12 considered animals, we recorded the activity of 2444 neurons (204 ± 69 per subject). We first asked how individual cells responded to tone presentations throughout the six days of the paradigm. Cells were classified as tone-responsive based on the average z-score of their trial-averaged tone response (example cells in Fig. 3.3A, see section 2.4.2 for classification procedure). Before conditioning only a small fraction (below 5%) of cells were classified as having a tone response (Fig. 3.3B). After conditioning, the fraction of CS+ responding cells, but not CS- responding cells, showed a marked increase, consistent with the behavioral discrimination between the two tones. The magnitude of post-conditioning tone responses peaked shortly after tone onset, but responses were sustained until the end of the tone (Fig. 3.3A,D). The population tone response that emerged with conditioning was based on cells gaining a positive or negative response to CS+ presentations rather than on previously tone-responsive cells strengthening their response (Fig. 3.3C,D shuffled cell IDs; Pearson correlation of 0.14 ± 0.08 between pre-conditioning and post-conditioning z-scores, mean \pm STD over 12 mice). Moreover, the distribution over all average z-scores became wider from pre- to post-conditioning (3.3E), indicating that the emerging tone representation was formed on the level of the population rather than by a small subset of cells. These results suggest that fear conditioning induces novel and distinct tone representations in mPFC population activity.

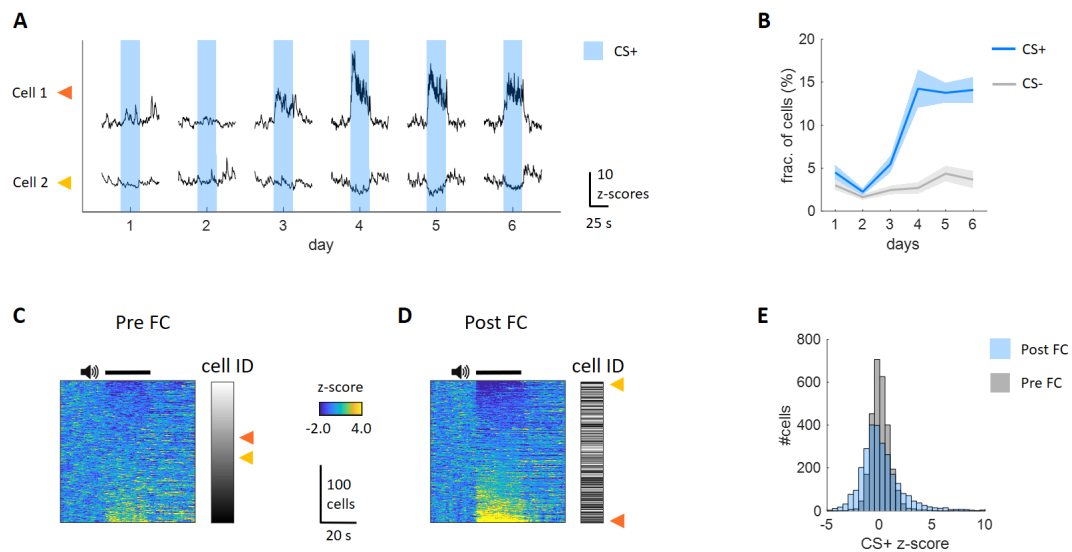


FIGURE 3.3: Single cell tone responses. **(A)** CS+ responses per day for two example cells. Tone-aligned activity was trial-averaged and scored. **(B)** Fraction of tone-responsive cells over days for CS+ and CS- (Mean \pm SEM, $n=12$ mice). To treat both tones equally, we only used the first four trials for trial averaging and z-score calculation. **(C)** Pre-conditioning CS+ responses of all cells of an example subject. Cells are sorted according to their mean z-score, cells IDs were assigned based on this sorting. Triangles indicate example cells from (A). **(D)** Post-conditioning analogue to (C). Cells were resorted according to post-conditioning mean z-score. **(E)** Distribution of mean z-scores over all animals for pre- and post-conditioning.

We next asked how the activity of individual cells related to freezing behavior. One issue with assessing freezing-related activity is that many cells in mPFC are modulated by general motion [77]. Since freezing is characterized by an absence of motion, motion-related activity is a clear confounding factor when analyzing the neural correlates of freezing. We found that many cells are clearly modulated by speed both in an excitatory and inhibitory way (example cells in Fig. 3.4A). To quantify how the activity of individual cells related to freezing and motion, we calculated freezing and speed scores per cell. The freezing score was defined as the correlation coefficient between the binary freezing trace and the activity of a given cell. Similarly, we defined a speed score as the correlation coefficient between neural activity and log-transformed speed (see 2.4.2). Importantly, the speed score was computed solely with data from outside freezing episodes in order to remove any correlation that might originate from a relation to freezing. We found that freezing and speed scores were highly correlated in all subjects and that there was no clear subpopulation of cells that was modulated by freezing, but not by general motion (Fig. 3.4B,D). While this analysis does not exclude the presence of freezing-specific activity, it shows the difficulty of analyzing behavioral signals in the presence of general motion correlates. Although we found clear correlates of freezing behavior, we did not find activity patterns that were clearly freezing-specific.

Nevertheless, we used the freezing score to ask if we could identify a connection between tone responses and freezing-related activity in single cells. Tone responses and freezing have a systematic temporal overlap (Fig. 3.2B), which contaminates tone z-scores with freezing related activity. To test whether there was joint coding between tone and freezing that goes beyond this temporal overlap we compared freezing scores to tone-onset z-scores, as the overlap with freezing is minimal at

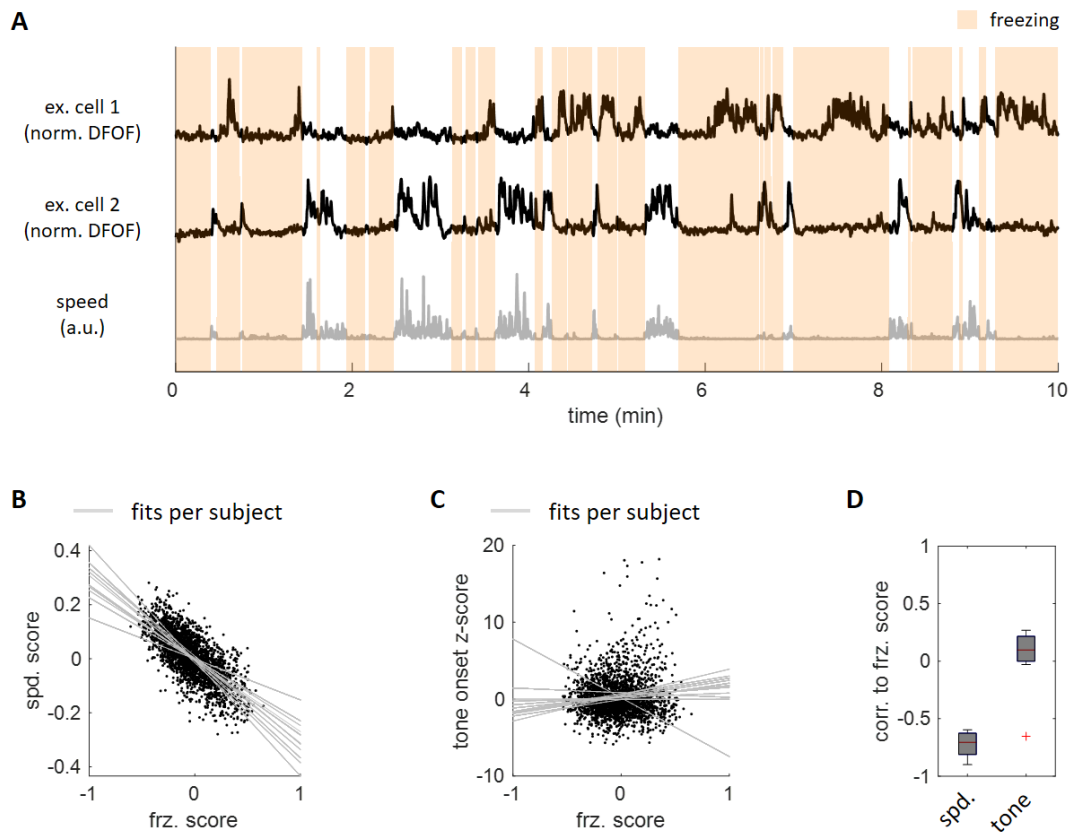


FIGURE 3.4: Single cell correlates of freezing. **(A)** Example cells showing increased and decreased activity during freezing periods and opposing modulation during periods of motion. **(B)** Relation of freezing and speed scores for individual cells. **(C)** Relation of freezing and tone-onset z-scores for individual cells. **(D)** Quantification of score correlations over 12 subjects.

tone onset. Based on this comparison we did not find any joint coding of tone and freezing information (Fig. 3.4C,D).

3.4 Population-level analysis of tone-evoked activity

To analyze tone responses and their evolution throughout the paradigm in more detail, we employed an approach based on decoding the presence of the tone stimulus from vectors of neural activity recorded at different time points. This approach enabled the flexible comparison of tone representations over different sessions, trials and time points during a trial. We used linear support vector machines (SVMs) to classify neural activity vectors for individual 200 ms time steps into either tone or baseline classes, indicating whether or not the CS+ was present at a given time step. As these classes are the same for all animals, we pooled cells from individual subjects to enable joint analysis and improve decoding performance (see section 2.4.3 for details on data structuring, training and testing procedures). Using this approach, we found that the presence of the CS+ could be reliably decoded after, but not before conditioning (Fig. 3.5A,B). CS- decoding accuracy also increased with conditioning, but remained substantially worse than CS+ decoding. Interestingly, decoding performance for the two tones closely resembled the evolution of the fraction of tone-responsive cells over days (Fig. 3.3B). To test how stable CS representations were over days, we evaluated how decoders trained on individual days

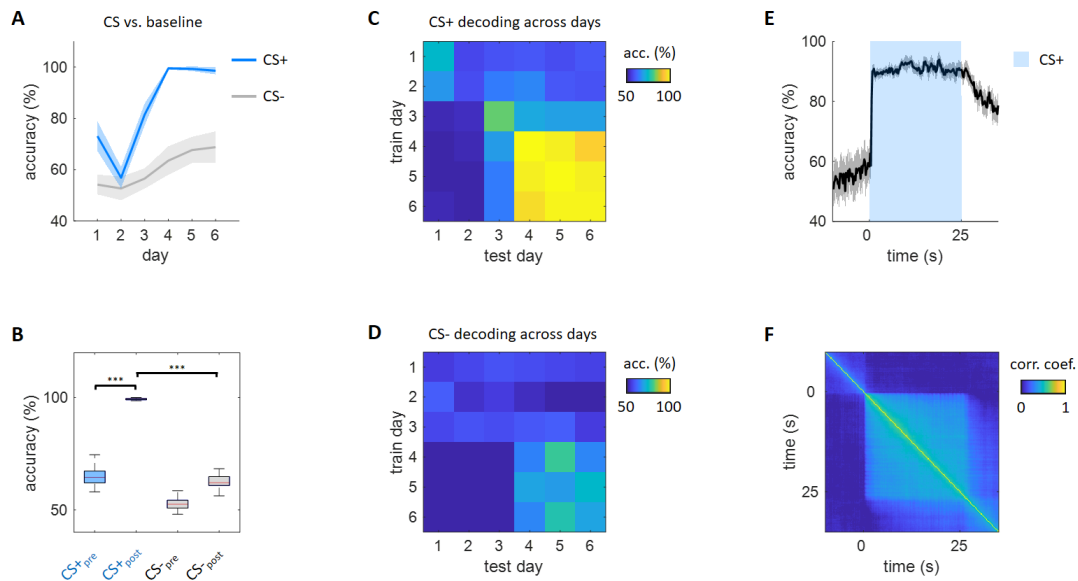


FIGURE 3.5: Tone coding over sessions. **(A)** Tone vs. baseline decoding over days. Mean \pm SD ($n = 50$ bootstrap runs). **(B)** Comparison of mean decoding accuracy for CS+ and CS- in pre- and post-conditioning sessions. Wilcoxon signed-rank test, $P < 0.001$. **(C)** CS+ vs. baseline decoding across days. Decoders were trained on one day and tested on all days. Color-code represents mean decoding accuracy over 50 bootstrap runs. **(D)** Same as (C) for CS-. **(E)** CS+ vs. baseline decoding per time step. For each time step, an individual decoder was trained on all post conditioning trials. Mean \pm SD ($n = 50$ bootstrap runs). **(F)** Similarity of decoders trained for different time steps, quantified by the pairwise correlation coefficient between decoder weight vectors.

performed when tested on data from the remaining days. We found that CS+ decoders trained on one of the post-conditioning days also showed good performance on other post-conditioning days, indicating that the tone responses emerging with fear conditioning stayed stable over days (Fig. 3.5C).

Next, we asked how CS+ responses evolved during the 25s time interval during which tones were presented. Instead of training one decoder for all time steps, we trained individual decoders with data from specific time points. Given the stability of tone responses over post-conditioning days, we pooled trials from these 3 days and trained one set of time-dependent decoders. We found that decoding accuracy was at a stably high level throughout the duration of the tone (Fig. 3.5E). After tone offset, accuracy did not immediately drop to chance levels, but slowly decayed. A possible explanation of this effect is that the freezing level remains elevated compared to baseline after CS+ presentations end (Fig. 3.2E), such that decoders could base the distinction on neural correlates of freezing behavior.

To assess whether the basis of the decoders' classifications changed over time, we compared the learned weights of SVMs trained on data from different time points. Decoders assign a weight to every cell, and weights can be used as a measure of how informative a cell's activity is to distinguish the two classes (tone and baseline). To quantify similarity between two decoders, we calculated the correlation coefficient between the decoders' weight vectors. Correlation coefficients between different time points stayed stably high, indicating that tone information was not represented by a dynamic code, but rather through the sustained responses of tone-responsive cells (Fig. 3.2F). The drop in correlation coefficient values after tone offset also supports the hypothesis that post-tone decoding performance is not based on

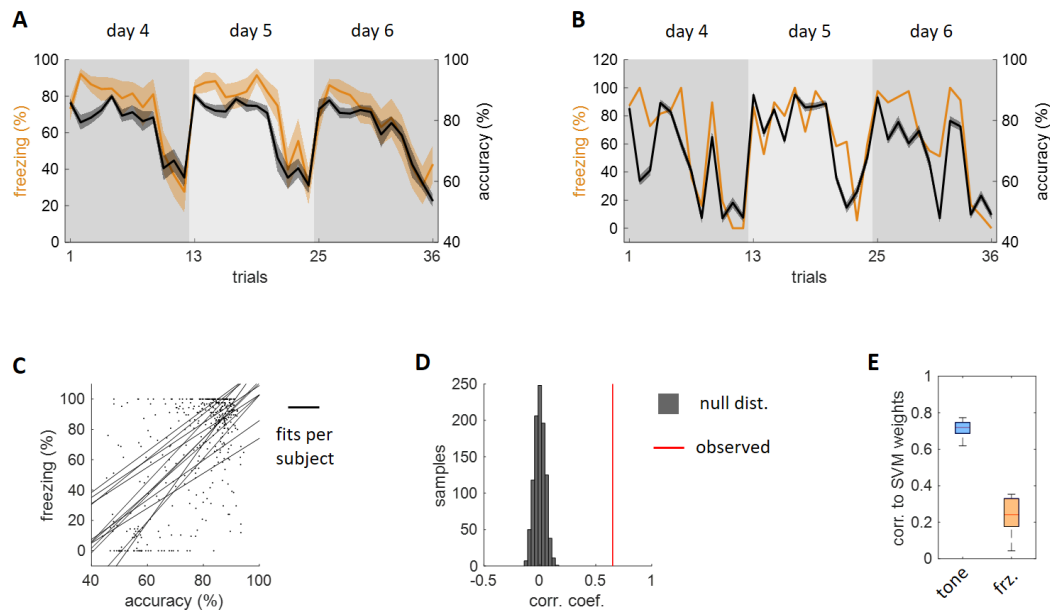


FIGURE 3.6: Tone responses correlate with freezing behavior. **(A)** Mean tone decoding accuracy and freezing strength over all post-conditioning trials. Mean \pm SEM ($n = 12$ subjects). **(B)** Same as **(A)** for an example subject. Mean \pm SEM ($n = 50$ bootstrap runs). **(C)** Relation of freezing strength and tone decoding accuracy over all trials. Lines represent correlation coefficients calculated per subject. **(D)** Permutation test comparing the observed mean correlation coefficient (red line) to hypothetical mean correlation coefficients obtained from 1000 repetitions with randomly shuffled data. **(E)** Correlation of decoder weights to tone z-scores and freezing scores. Decoder weights were obtained per subject by training on all 36 trials and calculating the average weights over 50 bootstrap iterations.

ongoing tone representations, but rather on other differences to the baseline state such as sustained freezing. Taken together these results show that fear conditioning induces tone-specific sensory responses that stay stable over days and are sustained throughout tone presentations.

3.5 Single-trial tone representations and their relation to freezing behavior

Since behavioral extinction had a clear temporal evolution over the trials of a session, we next asked if we could find similar dynamics in the evolution of tone representations over trials. To be able to relate tone information on individual trials to the behavior of individual subjects, we trained CS+ vs. baseline decoders per subject for all post-conditioning trials. For every given trial, we used all remaining trials as training data and quantified the accuracy of predicting the tone for the time steps of this trial. This approach revealed that tone decoding accuracy decayed towards the end of each session, mirroring the temporal dynamics of extinction as reflected by decreased freezing strength (Fig. 3.6A). The correlation between tone decoding accuracy and freezing strength was not only present when considering a mean over subjects, but also captured the variability for individual subjects (Fig. 3.6B,C). To assess the statistical significance of the observed correlation we performed a permutation test, which showed that the observed correlation coefficients were strongly increased as compared to shuffled data (Fig. 3.6D).

These results suggest that the presence of prefrontal tone responses is linked to the execution of freezing behavior. However, an alternative explanation could be that decoders base their classification not purely on tone-related activity. In particular, due to the increased freezing levels during the tone, freezing-related information could be used for the distinction between tone and baseline time steps in this decoding analysis, which could explain the reduced decoding accuracy for trials with low freezing. To investigate the basis of the decoders' classifications, we analyzed how the trained decoder weights related to tone and freezing scores. We found that decoding weights were strongly correlated to tone scores, but not freezing scores. These results indicate that the correlation between tone decoding accuracy and freezing behavior is indeed based on a lack of tone-related information in trials with low freezing.

In summary, we found that fear conditioning induces prefrontal responses to CS+ presentations whose presence correlates with the execution of freezing behavior. The identification of freezing-related activity was difficult because of confounding motion-related activity and we could not find a link between tone responses and neural correlates of freezing. Yet, the correlation between tone decoding accuracies and freezing behavior in single trials implies that there is some link between sensory-driven mPFC responses and behavioral execution. To better understand this link, we next designed aversive conditioning paradigms with changing stimulus-response mappings in order to be able to look for corresponding changes in the neural activity that might link stimuli and behavioral responses.

Chapter 4

Two-dimensional Active Avoidance

To further investigate the link between prefrontal tone representations and behavior execution, we developed a paradigm with a changing stimulus-response mapping. We designed a novel active avoidance experiment, in which we trained animals to adapt the instrumental action that was induced by presentations of an auditory cue in two consecutive learning tasks. This approach allowed us to investigate the link between the sensory stimulus and the behaviors it induced by analyzing changes in neural representations associated with the switch between the two instrumental behaviors learned in the two tasks of the paradigm. We start this chapter, by presenting the paradigm and characterizing animal behavior and learning performance. Subsequently, we analyze the relation of prefrontal neural activity to stimulus presentations and behavior execution and investigate learning-induced changes in these relations. Finally, we investigate how these sensory and behavioral representations are related.

4.1 Behavioral paradigm

The two-dimensional active avoidance paradigm had a duration of 11 days, comprising habituation (day 1), active avoidance task 1 (days 2-4), active avoidance task 2 (days 5-9) and extinction sessions (days 10 and 11). All sessions had a duration of

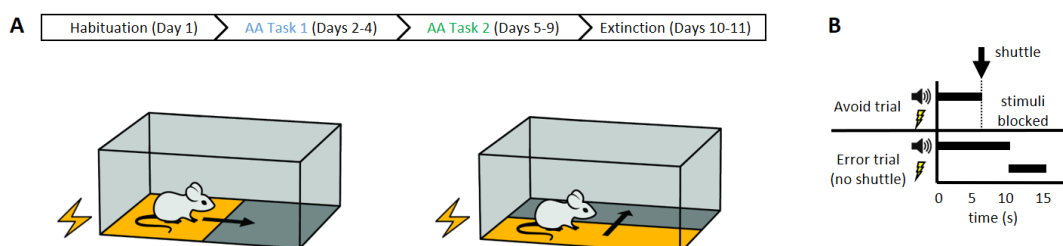


FIGURE 4.1: The 2-dimensional active avoidance paradigm (2DAA). **(A)** Schedule of the 2DAA paradigm. In task 1 (left) animals are trained to shuttle between the left and right compartments of the shuttle box in response to the tone. Which compartment receives the shock is determined by the position of the animal at trial start. In task 2 (right) the compartmentalization of the shuttle box is changed, and animals are trained to shuttle between the front and back half of the shuttle box. **(B)** Logic of active avoidance trials. If a subject performs the shuttle behavior before the end of the tone (10s), tone and shock presentations are blocked. If the subject does not shuttle, it receives a shock.

40 minutes and contained 50 trials with pseudorandom inter-trial intervals of 30 ± 10 s. Each trial started with the presentation of a 10 kHz tone at 75 dB for 10 seconds. In all active avoidance sessions (days 2 to 9) the tone was followed by a foot-shock (0.2 mA) with a maximum duration of 5 s. For each trial, we used automatically controlled, movable platforms (cf. section 2.2.4) to define half of the shuttle box as a safe zone. The position of the safe zone was determined by the task (1 or 2) and the position of the mouse at the start of a given trial. For task 1 trials, mice had to cross the vertical mid-line of the cage to reach the safe zone, whereas for task 2 trials, mice had to cross the horizontal mid-line (Fig. 4.1A). For the remainder of this thesis, these two actions will be referred as dimension 1 shuttling (D1-shuttling) and dimension 2 shuttling (D2-shuttling). If mice entered the safe zone by performing the appropriate shuttle behavior during tone or shock presentation, both tone and shock channels were blocked and the trial was terminated (Fig. 4.1B). If the behavior of mice did not lead to trial termination before shock onset, the two platforms in the safe zone were elevated for a duration of 15s, time-locked to the onset of the shock, providing mice with the possibility to escape. We recorded animal behavior over the full duration of the experiment using two top-view cameras.

4.2 Behavior quantification

We recorded data from 13 subjects, but had to exclude data from one animal due to low imaging quality which precluded the alignment of neural recordings between days. All subjects learned both tasks and avoided $84.3 \pm 2.1\%$ and $81.2 \pm 4.3\%$ of shocks in the final learning sessions of the tasks respectively (Fig. 4.2A). In general, performance increased within sessions, and the shuttle rate already substantially increased in the first learning session for both tasks. The task switch led to a clear drop in performance, but throughout task 2 sessions the shuttle rate continually increased and recovered to a level above 80% by day 9. As avoidance was based on two different actions in the two tasks, the task switch was reflected in the frequency of the two shuttle types during tone presentations on different days. While D1-shuttles were substantially more frequent than D2-shuttles in task 1, their frequency dropped in task 2 and D2-shuttle frequency increased (Fig. 4.2B). As intended by the experimental design, the two shuttle types were associated with different motion profiles (Fig. 4.2C,D). However there was considerable variability regarding the way animals adapted their behavior upon the task switch (Fig. 4.2E,F). In task 1, D1-shuttles were stereotyped and similar between subjects. In contrast, D2-shuttles in task 2 were more variable within each subject and different subjects crossed the mid-line in different ways. While some subjects used the shortest path to the other side (e.g. S1 in Fig. 4.2E), others shuttled diagonally. Yet, the consistent difference between shuttle angles for tasks 1 and 2 and the drop in performance induced by the task switch clearly indicate that mice had to change their behavior by learning to perform a different action in task 2.

We further analyzed avoidance actions by comparing the shuttle transitions that lead to avoidance to transitions during the inter trial interval (ITI). In the habituation session on day 1, the frequency of transitions was similar between time points during and outside the trial for both transition types (Fig. 4.3A,B). With the start of active avoidance conditioning on day 2, the frequency of ITI transitions decreased, while the frequency of transitions during tone presentations increased in a task-specific manner. In the extinction sessions on days 10 and 11 the frequency of transitions

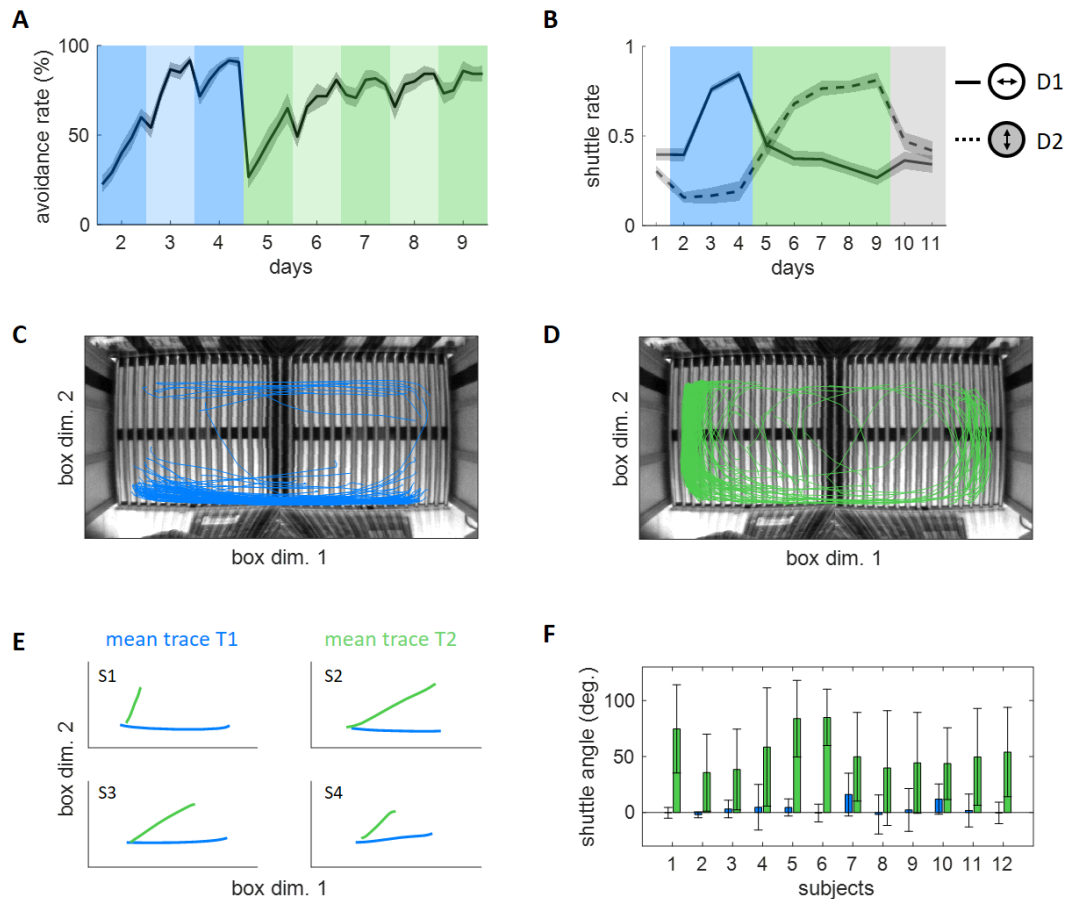


FIGURE 4.2: Evaluation of shuttle behaviors. **(A)** Fraction of avoid trials over task 1 (blue) and task 2 (green) sessions. Individual data points are an average over 10 trials, the displayed curve represents the mean \pm SEM over the 10-trial averages of 12 mice. **(B)** Frequency of the two different shuttle behaviors (D1-shuttling, solid line; D2-shuttling, dashed line) during tone presentation over days. Mean \pm SEM, $n=12$ mice. **(C)** Motion traces of task 1 avoidance shuttles of an example mouse overlaid on an image of the shuttle box. Motion traces were extracted for the windows ± 2 s around the detected shuttle time points. **(D)** Analogue of (C) for task 2 trials. **(E)** Mean motion traces for task 1 and 2 shuttle actions for four example mice. To compute the mean, all trials were adjusted such that the starting position of the mouse was in the lower left quadrant by mirroring along the appropriate axes. **(F)** Distribution of shuttle angles for tasks 1 and 2 over all animals.

during trials decreased, while ITI transitions frequencies increased, such that frequencies again had similar levels. These results further demonstrate that avoidance actions are a learned and tone-induced behavior.

Next, we compared tone and ITI transitions in terms of the average speed with which animals shuttled from one compartment to the other. The distributions of average speed values were overlapping, but showed a clear separation between avoidance and ITI transitions, where avoidance transitions were associated with higher speeds (Fig. 4.3C,D). This difference suggests a distinction between goal-directed avoidance actions and regular exploratory motion within the shuttle box. When comparing D1 with D2 transitions, both avoidance and ITI transitions had lower average speed for the latter. This difference can be explained by the available space for the shuttle motions. As the shuttle box is rectangular and animals spend most time in the corners, D1 transitions require them to cover greater distances, which

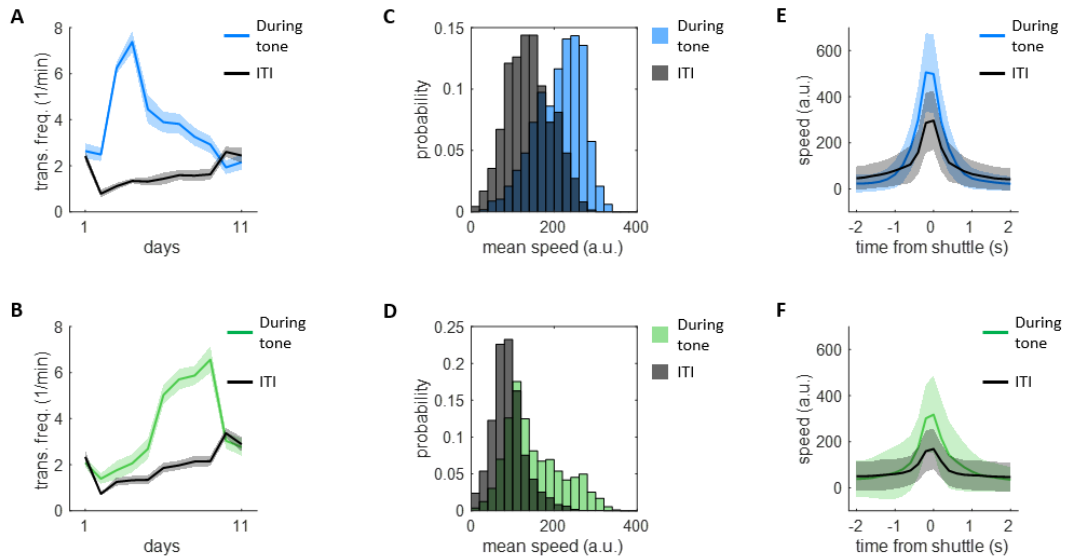


FIGURE 4.3: Comparison of shuttle behaviors and outside-trial transitions. **(A)** Frequency of left-right transitions during (blue) and outside (black) tone presentations. Mean \pm SEM over transitions. **(B)** Analogue to (A) for task 2 transitions. **(C)** Distributions of mean speed value during shuttling action for avoidance actions (blue) and outside-trial transitions (black). **(D)** Analogue to (C) for task 2. **(E)** Average speed during avoidance and outside-trial transitions. Mean \pm STD over transitions. **(F)** Analogue to (E) for task 2.

allows for higher acceleration (Fig. 4.2C,D). For all transitions, speed values were maximal around the time of the transition (Fig. 4.3E,F), but peak values were higher for avoidance actions, demonstrating that the difference in mean speed is based of a difference in maximal intensity. Taken together, these data show that avoidance actions differ from general motion, and that avoidance actions differed between the two tasks.

An additional way of characterizing learning-induced changes to animal behavior is the quantification of the latency of avoidance actions with respect to tone onset. We found that within tasks, latency values vary only slightly (Fig. 4.4A). Yet, there was a clear increase in latency of approximately 1s between task 1 and task 2. While task 1 latencies were approximately normally distributed around 5s, the distribution of task 2 latencies was skewed towards the end of the trial (Fig. 4.4B). Interestingly this increase in latency can be linked to the animals' freezing behavior. We observed

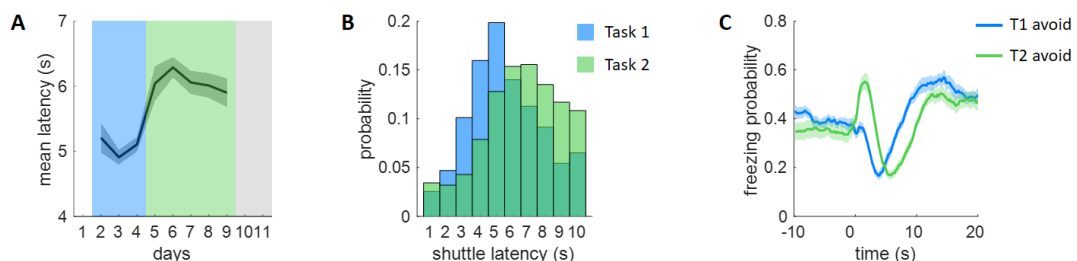


FIGURE 4.4: Latency of shuttle actions. **(A)** Mean shuttle latency over task 1 and task 2 sessions. Mean \pm SEM, $n=12$ mice. **(B)** Distribution of shuttle latency for task 1 (blue) and task 2 (green) avoidance shuttles. **(C)** Freezing probability over trial duration for task 1 and task 2 avoidance trials. Mean \pm SEM, $n=12$ mice.

that freezing levels were briefly elevated after tone onset for task 2, but not task 1, and that the freezing probability profile was shifted by one second. This observation could indicate a conflict between the execution of the two actions learned in tasks 1 and 2 that results in a brief startle reaction.

For the analysis of the neural processes leading up to avoidance it is crucial to accurately detect the start of avoidance actions. Typically, avoidance actions occur as a sudden onset of motion after a longer period of immobility. The identification of this action onset is particularly important for isolating avoidance-related activity that occurs before action onset from motion-related activity that occurs after action onset. For every avoidance trial, we use the speed of the animal to define the action start as the time point preceding the maximal increase in speed in a window of 2s before the transition was recorded (see Fig. 4.5A for example trials). For a majority of trials, the action start was determined to be approximately 1s before the transition was recorded and the tone was turned off (Fig. 4.5B). When using the tone end as an alignment point for averaging speed over avoidance trials, the mean speed peaks slightly before the alignment point (Fig. 4.5C). In contrast, using the action start alignment, speed values strongly increase and peak shortly after the alignment point. However, a slight motion increase before the action start alignment point remains, as avoidance actions do not always occur as a sharp transition from non-motion to motion. Nevertheless, this action onset alignment greatly facilitates the analysis of neural processes that precede the execution of avoidance actions.

4.3 Imaging data set

Throughout the 2DAA paradigm, we recorded the activity of 3326 cells from 12 mice (277 ± 51 cells per subject). For 3 subjects we had to exclude data from the last one or two days (extinction sessions) due to alignment issues. For all other subjects, however, we could follow the activity of all cells over all sessions. Fig. 4.6A shows the spontaneous activity of 10 example cells recorded on different days. Relating the recorded neural activity to experimental variables is generally challenging, and our setting combines several issues that make it difficult to directly interpret the recorded activity with respect to our experiment. First, there is an inherent overlap between several experimental variables, in particular between action related activity and tone-related activity and between action-related and motion-related activity, as illustrated in Fig. 4.6B. Second, both stimulus presentations and behavior are variable over trials, precluding straight-forward trial averaging approaches. Third, the

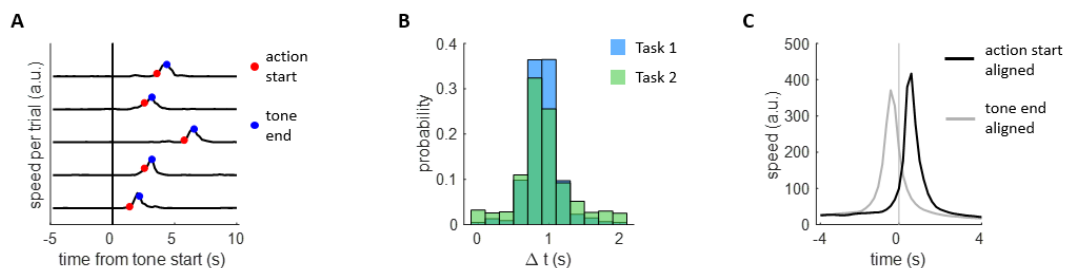


FIGURE 4.5: Distinction of shuttle time and action start time. **(A)** Speed traces of five example cells with shuttle time (tone end) and action start time indicated. **(B)** Distribution of temporal distance between action start and tone end. **(C)** Mean speed over all avoidance trials aligned to action start (black) or tone end (gray).

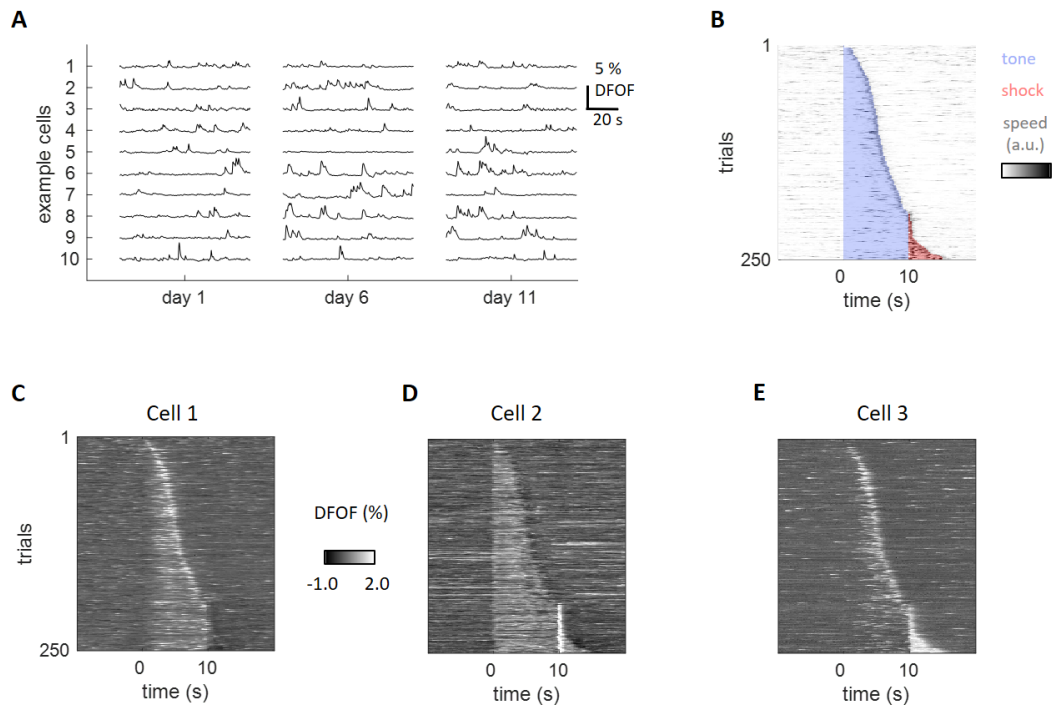


FIGURE 4.6: Example cell activity. **(A)** Spontaneous activity for 10 example cells across the 11-day paradigm. **(B)** Illustration of the variability and overlap of experimental variables over trials for an example subject. Trials are sorted according to tone length. **(C)** Example cell whose activity correlates with tone presentations and avoidance actions. **(D)** Example cell whose activity correlates with tone and shock presentations. **(E)** Example cell whose activity correlates with motion.

activity of the recorded cells is often correlated with multiple experimental variables as displayed in the example cells in Fig. 4.6C-E. Fourth, we want to understand the dynamics unfolding over the course of tone presentations rather than static stimulus representations. And finally, we want to assess changes in coding and dynamics over time, and in particular in relation to the switch between the two tasks of the 2DAA experiment. However, for all of these challenges, there exist approaches that allow addressing them. In the following sections we use these approaches to investigate the neural correlates of tones and avoidance actions. We start by investigating how stimuli and behaviors are represented at the level of single cells and neural populations, and analyze how these representations change over time and upon the task switch. Based on these representations we then address the question of whether we can find a systematic link between tone-evoked responses and the execution of avoidance actions in the recorded neural activity.

4.4 Tone-evoked activity

We found that many cells show temporally precise responses to tone onset and offset (Fig. 4.7A,C). To quantify tone responses, we calculated a trial average, where for every time step we only averaged over trials in which tone presentations were still ongoing, in order to address the variability in tone length over trials (Fig. 4.7B,D). Based on these trial averages, we calculated a z-score using pre-tone activity as a baseline, and classified cells as tone-responsive, if the absolute value of the average

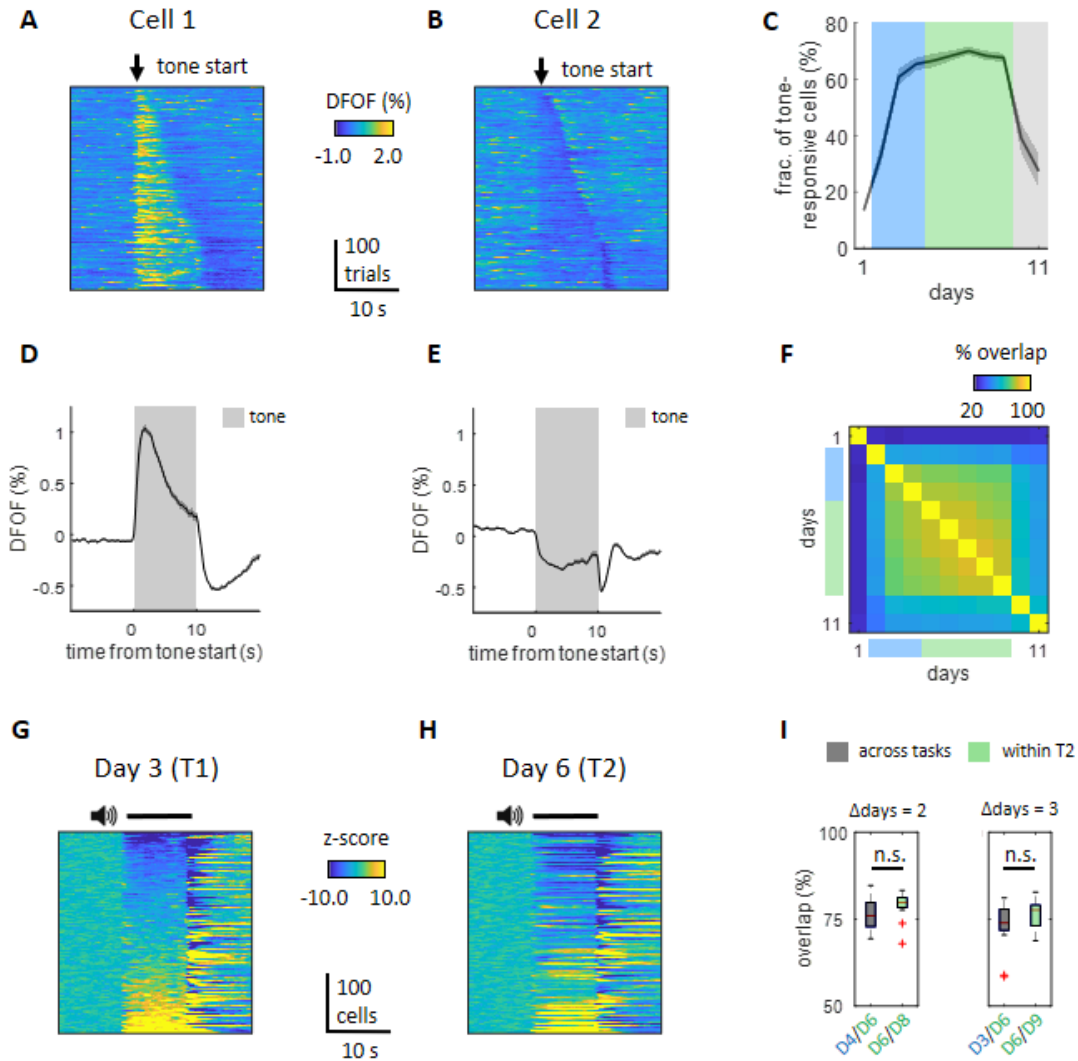


FIGURE 4.7: Single cell tone responses. **(A,B)** Example cells with positive and negative tone responses. Trials are aligned at tone start and sorted according to tone end. **(C)** Fraction of tone-responsive cells over sessions. Mean \pm SEM, $n=12$ mice. **(D,E)** Trial-averaged tone responses of cells from **(A)** and **(B)**. Time points after tone end are excluded from averaging for avoidance trials. **(F)** Overlap of tone-responsive cells for pairwise day comparisons. **(G)** Trial-averaged tone responses of all cells of an example subject on day 3. Cells are sorted according to their mean z-score during the tone. **(H)** Trial-averaged tone responses of all cells of the example subject from **(G)** on day 3. Cells are sorted as in **(G)**, highlighting the similarity in tone responses over days. **(I)** Quantification of task-specificity of tone responses comparing within-task and across-task similarity at two different time shifts. Wilcoxon signed rank test ($p>0.05$), $n=12$ subjects.

z-score over the full tone presentation exceeded a cutoff value of 2. Using this quantification, we found that the fraction of tone-responsive cells substantially increased during and after the first active avoidance session (Fig. 4.7C, increases of 20.1% and 27.2% respectively). For the rest of avoidance sessions, the fraction of tone-responsive cells remained relatively constant at a value of $66.8\% \pm 1.5\%$. In extinction sessions we observed a substantial drop in the number of tone-responsive cells, highlighting the tight connection between prefrontal sensory-evoked responses and behavioral relevance.

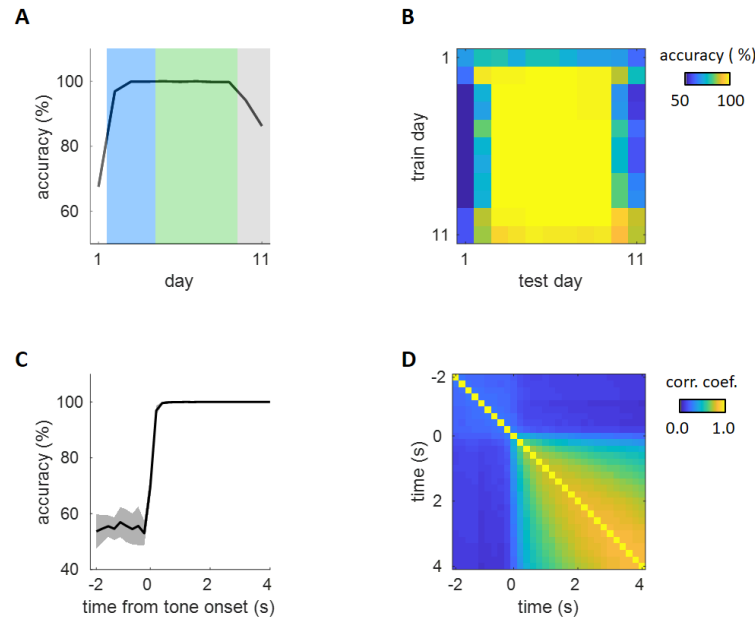


FIGURE 4.8: Decoding of tone information. **(A)** Tone vs. baseline decoding with individual decoders trained per day. Mean \pm SEM, $n=50$ bootstrap runs. **(B)** Across day testing for decoders from (A). **(C)** Time-resolved decoding of tone information vs. baseline. Mean \pm SEM, $n=20$ bootstrap runs. **(D)** Pairwise comparison of SVM weight vectors between all time steps from (C), indicating changes in tone representations over time.

As in our fear conditioning data, tone responses were both positive and negative and many cells maintained their response profile over multiple days (Fig. 4.7G,H). To quantify the stability of tone representations over days, we calculated the pairwise overlap between the subsets of tone-responsive cells over all days (Fig. 4.7F). The overlap was elevated for comparisons between days 3 to 9, reflecting increased similarity of tone responses from the day after the first conditioning session to the last conditioning session. To assess whether the task switch had an effect on tone coding, we compared the similarity of tone representations within and between tasks at multiple time lags (Fig. 4.7F). Using this approach, we did not find that the task switch induced a reorganization of tone responses. This suggests that mPFC tone responses signal behavioral relevance without capturing aspects of the associated behavior, rather than representing tones in a context-specific manner.

To analyze tone responses in more detail, we used SVM classifiers to investigate how tone information was represented on the population level on different days and at different time points in a trial. As in the analysis of our fear conditioning data, we pooled cells over subjects and trained decoders to classify activity vectors of individual 5 Hz time steps as tone or baseline. First, we trained individual decoders for all 11 recording days. Consistent with our single cell results, tone decoding worked very well in active avoidance sessions, but performance was worse in habituation and extinction sessions (Fig. 4.8A). When testing individual per-day decoders on other days, we found that across-day decoding worked well within, but not outside active avoidance sessions, indicating consistency of the tone representation during conditioning sessions (Fig. 4.8B). The first learning session on day 2 displays an interesting exception: the decoder trained on day 2 works well on the following days, however, data from data 2 is not correctly classified by decoders from days 3 to 9. This discrepancy indicates that the tone representations present on later days are

formed on day 2, but are only weakly present such that decoders from days 3 to 9 do not recognize them in day 2 data. Notably, the consistent performance over all active avoidance sessions again does not indicate any task specificity of tone representations.

We next asked if we could find changes in tone representations over the duration of the trial window. As behavior execution towards the end of the trial window provides a confounding distinction between tone and baseline time steps, we limited this analysis to the first 4 seconds of the trial and excluded trials where avoidance actions were performed before this cutoff. Since previous analysis indicated that there is no clear difference in tone coding between tasks, we trained per time step decoders using all trials from days 3 to 9. As expected from the performance of per day decoders, tone vs. baseline classes were perfectly after tone onset (Fig. 4.8C). However, the predictive information might be present in different cells at different time steps and we can use the weights of the individually trained decoders to analyze whether there was a change in how cells contribute to the classification decision. Across-time step weight comparisons revealed that tone coding was dynamic within the first 2s of tone presentations, indicated by decreased similarity of weight vectors when moving away from the diagonal (Fig. 4.8D). After approximately 2s, decoders became more similar, suggesting that tone responses settled into a steady representation after the initial dynamic phase. These results demonstrate that prefrontal tone responses are stable across both trial time and days. This stability is consistent with the interpretation of mPFC tone responses as a threat signal that stays constant over the experiment.

4.5 Avoidance-related activity

The identification of neural correlates of avoidance actions is challenging, as they overlap with other drivers of neural activity. First, avoidance-related activity has to be distinguished from general motion-related activity. Second, by design, avoidance actions can only occur during tone presentations, such that activity related to avoidance needs to be isolated from general tone-evoked activity. However, both of these issues can be addressed with appropriate comparisons. We approach these issues by analyzing activity before motion onset and by isolating avoidance-related activity from tone-related activity through a comparison of avoidance and error trials.

Any cognitive processes that precede the execution of goal-directed avoidance actions need to occur during the period shortly before action start. We used the action start time points as displayed in Fig. 4.5 to align all avoidance trials in order to facilitate the identification of pre-avoidance activity. However, as mentioned above, we need to take into account concurrent tone-related activity. We can isolate avoidance-related activity from tone-related activity using a comparison to error trials, as the two trial types have the presence of the tone in common, but only avoidance trials contain activity related to upcoming avoidance actions. Since tone presentations are shorter in avoidance trials, we aligned error trials such that their duration matches the average alignment time of avoidance trials to account for variability in tone-evoked activity over the trial duration. These two alignment strategies for avoidance and error trials are displayed for 4 example cells in Fig. 4.9A. Based on the aligned trials, we calculated trial averages for the two trial groups and defined a discrimination score that reflects how well a cell's activity can be used to distinguish between

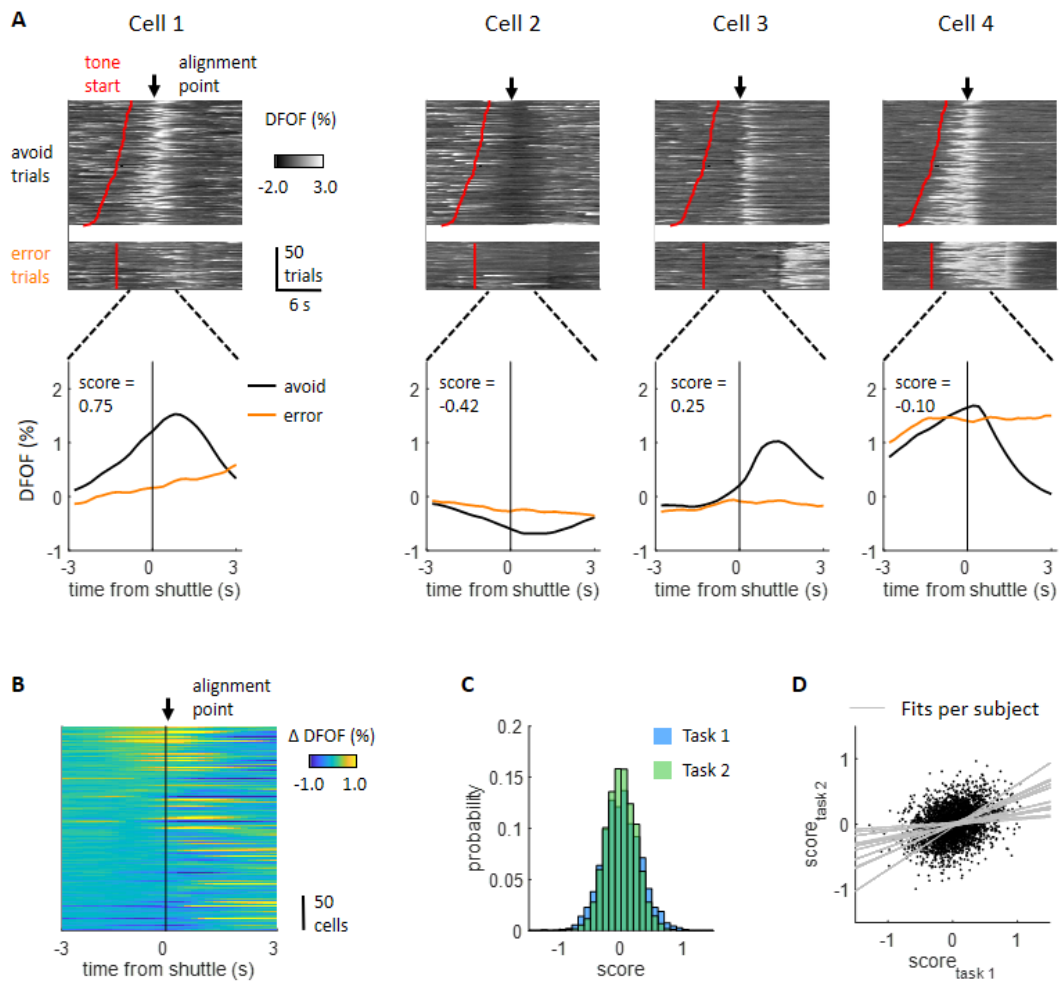


FIGURE 4.9: Single cell correlates of avoidance actions. **(A)** Four example cells showing different relations to avoidance actions. (Top row) Neural activity during avoidance and error trials. Avoidance trials are aligned according to action starts. Error trials are aligned to match the mean tone duration of avoidance trials. (Bottom row) Trial averages for aligned avoidance and error trials and avoidance score calculated for the individual cells. **(B)** Difference between avoidance and error trial averages for an example subject. Avoidance scores are computed based on this difference (see methods) and cells are sorted by avoidance score. **(C)** Distribution of avoidance scores over all cells for tasks 1 and 2. **(D)** Correlation of avoidance scores between tasks.

avoidance and error trials before action start (see section 2.4.2). Positive scores indicate that activity in avoidance trials is elevated in comparison to error trials, while negative scores indicate decreased activity. Cells 1 and 2 in Fig. 4.9A are examples of cells that distinguish between avoidance and error trials before action onset. In contrast, cell 3 only shows a clear separation between the trial types after action onset, which can be attributed to motion. Cell 4 has a clear tone response, which however does not show substantial differences between avoid and error trials, highlighting the importance of the comparison.

We found that cells which distinguished trial types before action onset usually also did so after action onset (Fig. 4.9B). In contrast, many cells that showed clear differences after action onset, did not differ between trial types before action onset (Fig. 4.9B). This contrast indicates a distinction between general motion coding

and preparatory activity related to goal-directed avoidance actions. To assess task-related changes in avoidance activity, we calculated two avoidance scores for every cell based on trials from the two tasks. For both tasks the avoidance score was approximately normally distributed and the distribution did not substantially differ between tasks (Fig. 4.9C). We next investigated the correlation between task 1 and task 2 avoidance scores and found a slight positive correlation for all subjects (4.9D). However, since avoidance scores for individual cells depended on trial averaging and were generally noisy, we chose to base further investigations of avoidance-predictive activity on analyses based on SVM decoders.

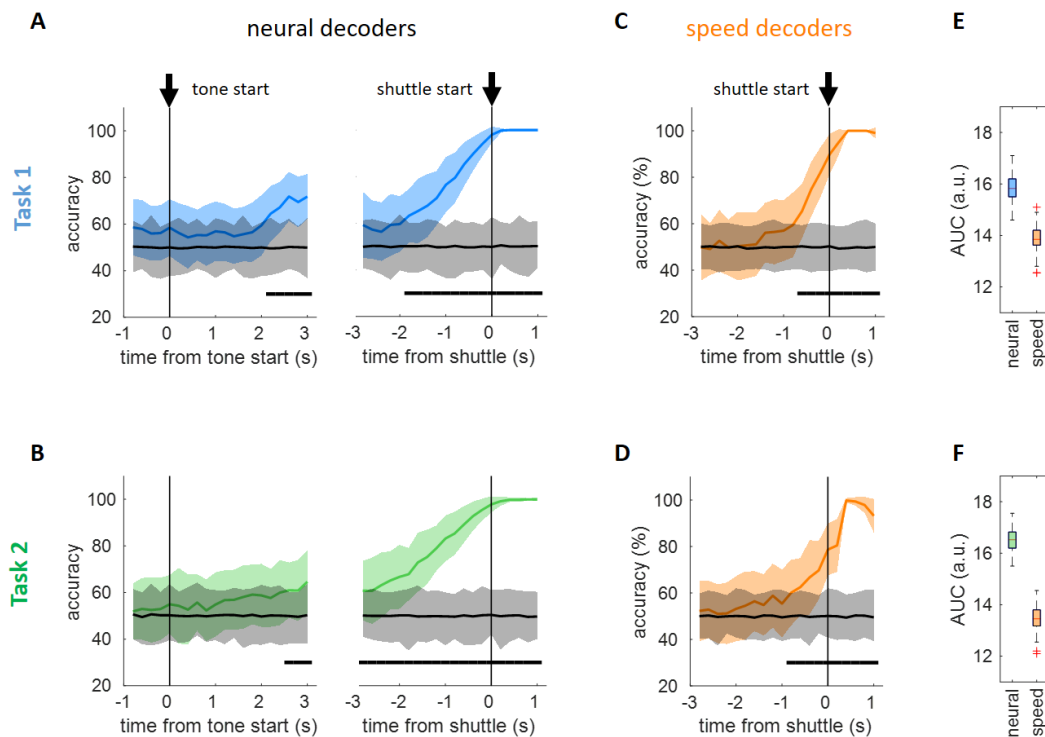


FIGURE 4.10: Prediction of avoidance actions. **(A)** Time-resolved decoding of trial-type (avoid or error) for time steps aligned to tone start (left) or shuttle action start (right). Decoding performance (blue) is displayed as mean \pm SD over 100 bootstrap runs. Black line and gray area represent mean and 0.5th and 99.5th percentiles of the distribution of mean accuracies of 200 shuffle repetitions of a permutation test (see methods). Black line at the bottom indicates significance ($p < 0.01$). **(B)** Same as (A) for task 2. **(C)** Trial type decoding based on motion information (same as right plot in (A) otherwise). **(D)** Same as (C) for task 2. **(E)** Quantification of decoding performance over 100 bootstrap runs. Area under the curve (AUC) is calculated over the whole period from 3s before to 1s after action start. **(F)** Same as (E) for task 2.

The decoding-based analysis of avoidance-predictive activity has the advantages that we can integrate information over many noisy cells and that we can assess predictive information in a time-resolved manner. We pooled cell activities over subjects and trials and trained SVMs to distinguish avoidance from error trials at different time points during the trial (200ms bins). Given the temporal variability over trials, we used two different alignment conditions: alignment to tone start and alignment to action start. In order to assess potential differences between avoidance actions in tasks 1 and 2, we trained separate groups of task-specific decoders. Detailed data structuring procedures are described in section 2.4.3. At tone start, decoding accuracy did not exceed chance level (Fig. 4.10A,B left). However, 2-3

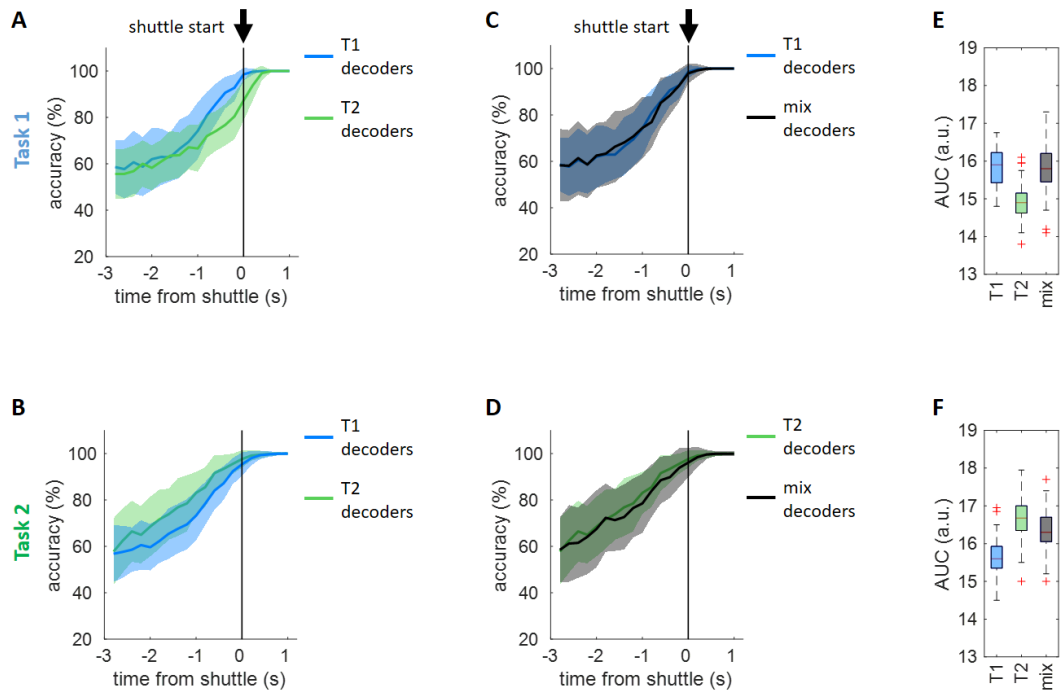


FIGURE 4.11: Action prediction performance evaluated across tasks. **(A)** Prediction of trial types for task 1 trials, using decoders trained on task 1 (within task testing) or trained on task 2 (across task testing). **(B)** Same as (A) for task 2 trials. **(C)** Comparison of within-task performance with decoders trained using data from both tasks (mix decoders). **(D)** Same as (C) for task 2 trials, mix decoders are the same for (C) and (D). **(E,F)** Quantification of decoder performance for the 3 decoder sets ($n = 100$ bootstrap runs).

seconds after tone onset, we observed an increase of decoding accuracy for both tasks. The action start alignment condition revealed that upcoming avoidance actions could be predicted with above-chance accuracy levels already 3 seconds before action onset (Fig. 4.10A,B right). Decoding accuracy increased up to action onset and actions were perfectly decoded after action onset, as expected due to the presence of motion related activity in avoidance, but not error trials. Since action start alignment can not entirely exclude pre-action motion, we next asked whether the observed decoding performance could be reproduced by a control decoder using tracking-based motion information to predict upcoming avoidance actions. We found that motion decoders could also successfully decode actions already before action onset, however at a markedly delayed time as compared to neural decoders (Fig. 4.10C-F). This difference suggests that there indeed exists neural activity predicting future avoidance actions and indicates that the activity of prefrontal neurons reflects a cognitive process preceding the execution of avoidance actions.

The comparison of task 1 and task 2 decoders in Fig. 4.10 shows that in task 2, upcoming actions can be predicted slightly earlier, but overall the results were similar. To analyze if the neural activity predicting avoidance actions changed between tasks we further compared the two sets of decoders trained on data from the two different tasks. In particular, we asked how decoders trained on one task performed on data from the other task. Across-task testing showed drops in performance for both tasks (Fig. 4.11A,B). However, this drop does not necessarily indicate a change in action predicting activity, but might rather be a reflection of more general changes in

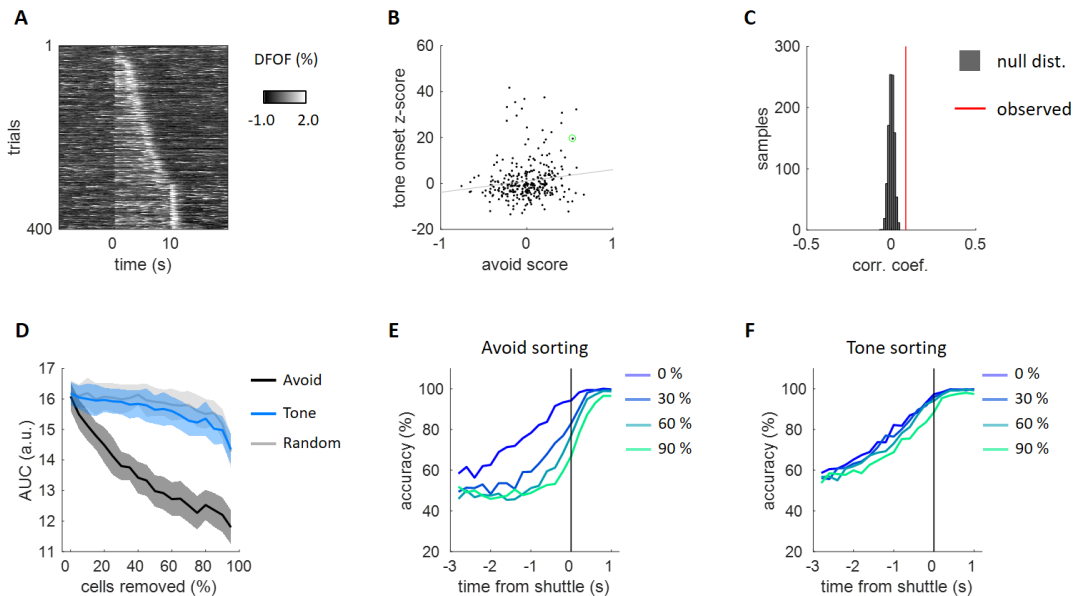


FIGURE 4.12: Relation of tone-evoked and avoidance-related activity. **(A)** Example cell displaying a tone-onset response and containing action-predictive information (activity increase around avoidance actions). **(B)** Correlation of tone-onset z-score and avoidance score. Example cell from **(A)** is marked in green. **(C)** Comparison of the observed mean correlation coefficient over subjects to correlation coefficients calculated from shuffled data. **(D)** Dependence of avoidance prediction performance on different cell groups. Cells are removed either according to avoidance score sorting (black), tone-onset z-score sorting (blue), or randomly (gray). **(E)** Decoding performance curves for different cell subsets; removal based on avoidance score sorting. **(F)** Same as **(E)** for tone-onset z-score sorting.

neural activity between tasks. To address this alternative explanation, we trained decoders mixing data from both tasks and compared their performance to within-task decoders (Fig. 4.11C,D). The set of mix decoders closely matched the performance of both task-specific decoders. As one set of mix decoders was able to reproduce the performance of two sets of task-specific decoders, these results indicate that predictive information had similar structure in the two tasks (Fig. 4.11E,F). Thus, the differences observed in across task decoding do not seem to originate from task-specificity of action-predictive activity. Taken together, these decoding results show that mPFC activity does contain action predictive information, however upcoming actions are not represented in an action-specific manner.

4.6 Relating tone and avoidance coding

Having analyzed the neural correlates of tone and avoidance actions, we next asked how these correlates are related. We first considered the avoidance and tone scores that we defined for individual cells and asked how they were related. As the tone-z score might also capture avoidance-related activity due to the systematic overlap of tone and avoidance, we calculated a tone-onset z-score based on the first 2s of the tone response. Using the tone and avoidance scores, we could identify individual cells that both responded to the tone and contained avoidance predictive information (Fig. 4.12A). On the level of the population however, there was no coordinated relation between the two types of information as captured by the scores (Fig. 4.12B,C). Yet, as the avoidance score relied on trial averaging and as pre-action

differences between avoid and error trials were generally small, it could be that the score did not accurately reflect avoidance-related information.

Since population-based decoding provided a better handle on avoidance-related information, we next devised an approach to analyze the importance of different subsets of cells with respect to decoding accuracy. To measure the importance of a subset of cells, we reran the avoid vs. error trial decoding, but removed the cell subset from the feature set normally consisting of all cells. We then quantified the change in decoding performance using the area under the curve as a summary statistic for the decoders trained on different time steps. We first analyzed how decoding performance evolved when removing cells according to their avoidance score. For this we sorted cells based on the absolute value of their avoidance score in descending order and then removed cell subsets of increasing sizes and quantified the resulting decoding performance (Fig. 4.12D). We observed that performance rapidly decreased upon the removal of even small subsets of cells, indicating that the avoidance score indeed captured the predictive information used by SVM classifiers. Removal of increasingly bigger subsets progressively delayed the decoding performance before action start, while decoding after action start was less affected (Fig. 4.12E). In contrast, removal of cells according to their tone-onset z-score had little effect on decoding performance and closely resembled the removal of random cells (Fig. 4.12D,F). These results are consistent with the analysis of single cell avoidance and tone scores and further indicate that there was no coordinated coding of tone and avoidance related information. We could thus not identify a transformation of tone-related information into action-related signals based on these analyses.

In summary, we established a novel active avoidance paradigm with a changing stimulus-response mapping that allowed us to investigate how mPFC activity relates to tones, actions and the association of sensory signals with behavioral responses. We found pronounced tone responses that depended on behavioral relevance, remained stable over conditioning sessions and did not show clear changes upon the switch of the tone-induced action with the relearning between tasks 1 and 2. Neural correlates of avoidance actions were more subtle, but we found that mPFC population activity contained predictive information about upcoming avoidance actions up to 3s before action start. However these predictive signals were not specific to the executed action and we could not identify a clear relation between the neural correlates of tone and avoidance.

Chapter 5

Two-tone Active Avoidance

The 2DAA experiment allowed us to investigate changes in stimulus-response mappings on the behavioral side. In a complementary experiment we trained mice to perform a single avoidance behavior in response to two different tones in two consecutive tasks. This setting allowed us to ask how prefrontal tone responses evolve as their relation to behavior changes. As the behavioral paradigm is closely related to the 2DAA setting, we use similar quantification methods and highlight similarities and differences between the two paradigms. We again start by introducing and quantifying the learning behavior and then quantify neural activity in terms of its relation to tones and avoidance actions.

5.1 Behavioral paradigm

The two-tone active avoidance (2TAA) paradigm consisted of 9 sessions comprising habituation (day 1), active avoidance task 1 (days 2-4), active avoidance task 2 (days 5-7) and extinction sessions (days 8 and 9). Sessions and trials had the same structure as for the 2DAA experiment (see section 4.1), with the exception that the 50 trials were randomly split into CS1 and CS2 trials. We used 4 and 10 kHz tones played at 75 dB and counterbalanced their assignment to CS1 and CS2 across animals. The pairing of tone and shock was dependent on tone type and task type. In task 1, CS1 was followed by a shock, while CS2 presentations did not have any consequence. In task 2, these contingencies were reversed. We activated shocks only in the compartment in which a subject was located at the beginning of a trial, leaving the other compartment as a safe zone. If mice shuttled between compartments during shock-paired trials, we blocked both tone and shock channels until the end of the trial (Fig. 5.1). As for the 2DAA experiment we recorded all behavior using two top-view video cameras.

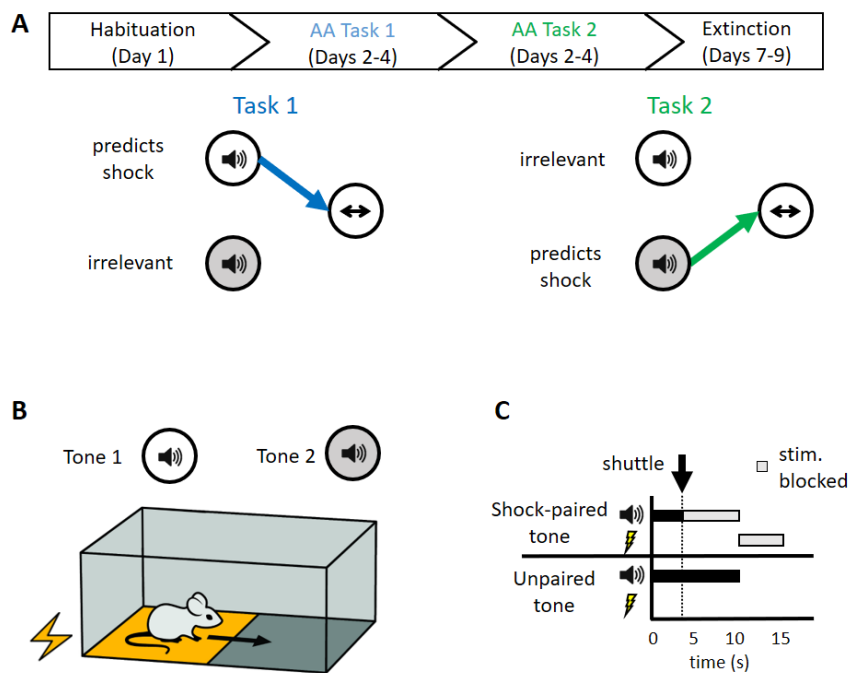


FIGURE 5.1: 2TAA paradigm. **(A)** Schedule of the 2-tone active avoidance paradigm. In task 1 (left) animals are trained to shuttle in response to tone 1 through the association to the shock, while tone 2 does not have any consequences. In task 2 (right) this mapping is reversed. **(B)** Schematic of shuttle box. **(C)** Logic of active avoidance trials. For shock paired trials, shuttling immediately blocks both tone and shock presentations. For tone presentations that are not paired with shock presentations, shuttling does not have any effect and the tone is always played for the full 10s.

5.2 Behavior quantification

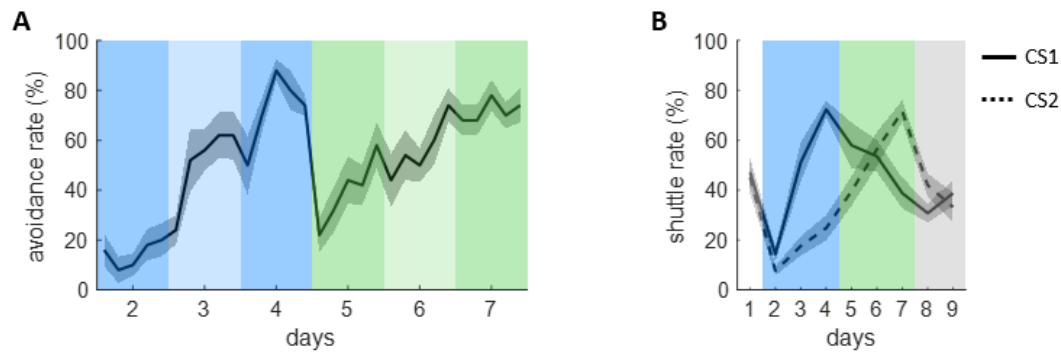


FIGURE 5.2: Analysis of shuttling behavior. **(A)** Fraction of avoidance trials of task 1 (blue) and task 2 (green) sessions. Individual data points represent and average over 5 trials and data is displayed as mean \pm SEM over 10 mice. **(B)** Frequency of shuttle action during CS1 (solid line) and CS2 (dashed line) presentations. Mean \pm SEM, $n=10$ mice.

We trained 16 animals on the 2TAA task, however 6 animals did not learn task 2 (data not shown) and were excluded from further analysis. The remaining 10 animals learned both tasks and on average reached performance levels of over 70% by the end of the last learning session (Fig. 5.2A). The task switch lead to a clear drop in performance and animals relearned to avoid shocks at a similar speed as for the learning of the first avoidance action. In comparison to the 2DAA experiment, learning was slower, which can however be explained by the lower number of trials per session (25 shocked tones vs. 50 shocked tones). In both tasks of the 2TAA experiment, animals showed behavioral discrimination between the two tones (Fig. 5.2B). In task 1, CS2, which did not have a consequence, caused only low shuttling rates. After the task switch, the frequency of CS2-induced shuttling continually increased, while CS1-induced shuttling continually decreased, due to the loss of CS1's connection to the shock. These results show that animals learned the mapping between the different tones and the shuttling behavior in a task-specific manner.

We next compared avoidance actions to transitions during the inter-trial interval. As in the 2DAA experiment, the frequency of ITI transition dropped with the onset of conditioning on day 2, while shuttling during tone presentations was increased in a tone-specific and task-specific manner (Fig. 5.3A). When analyzing the mean speed of the recorded transitions, we found that not only the frequency, but also the intensity of shuttle behaviors during the two tones changed between tasks (Fig. 5.3B), further demonstrating the adaptation of tone-induced behaviors. In general, shuttling during the tone was associated with higher speeds than shuttles during the ITI (Fig. 5.3C,D) and speeds peaked at the time of the shuttle (Fig. 5.3E,F).

Shuttle latency had comparable levels to the 2DAA experiment, and the mean shuttle latency also increased upon the task switch (Fig. 5.3G,H). However, as opposed to the 2DAA experiment, this increase was not connected to an increase in freezing probability at tone onset (Fig. 5.3I).

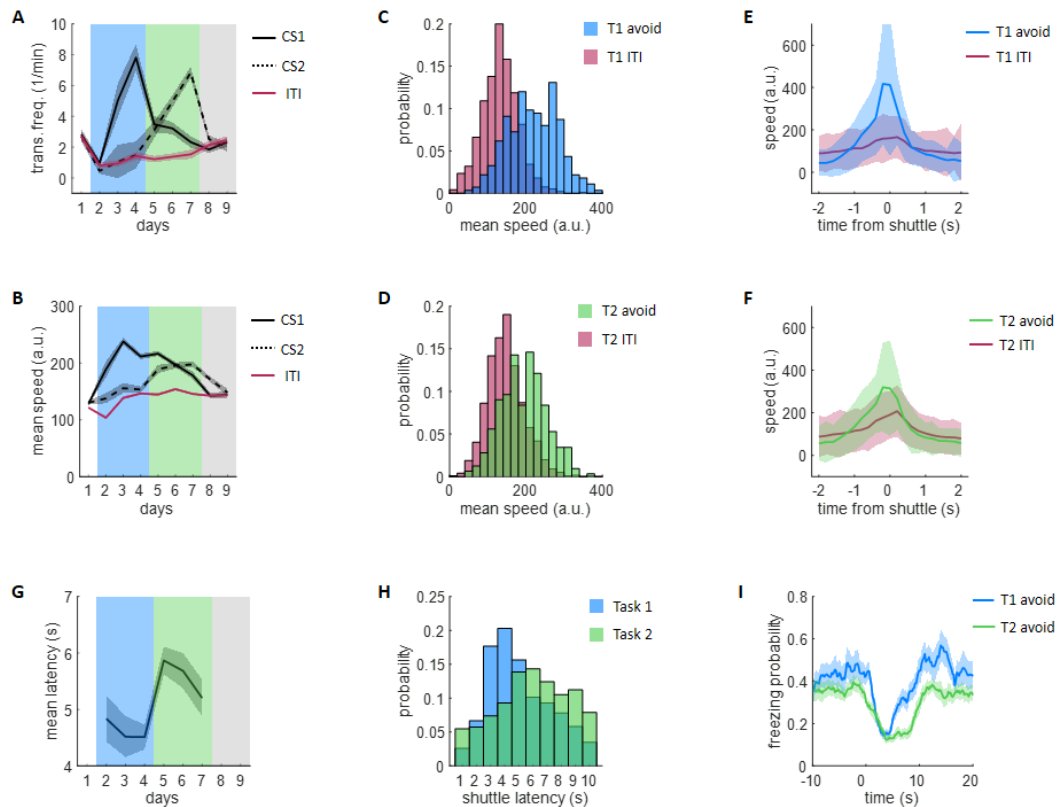


FIGURE 5.3: Comparison of avoidance actions and inter-trial interval (ITI) shuttles and shuttle latency. **(A)** Frequency of shuttling during tones and during the inter-trial interval. Same data as Fig. 5.2, but normalized by tone time. Mean \pm SEM, $n=10$ mice. **(B)** Mean speed of CS1, CS2 and ITI shuttles. Mean \pm SEM, $n=10$ mice. **(C)** Distribution of mean speed of task 1 avoidance and task 1 ITI shuttles. **(D)** Analogue of (B) for task 2 **(E)** Speed during task 1 avoidance action and task 1 ITI shuttles. Mean \pm STD over transitions. **(F)** Analogue of (E) for task 2. **(G)** Mean shuttle latency over task 1 and task 2 sessions. Mean \pm SEM, $n=10$ mice. **(H)** Distribution of shuttle latency for tasks 1 and 2. **(I)** Freezing probability during task 1 and task 2 avoidance trials. Mean \pm SEM, $n=10$

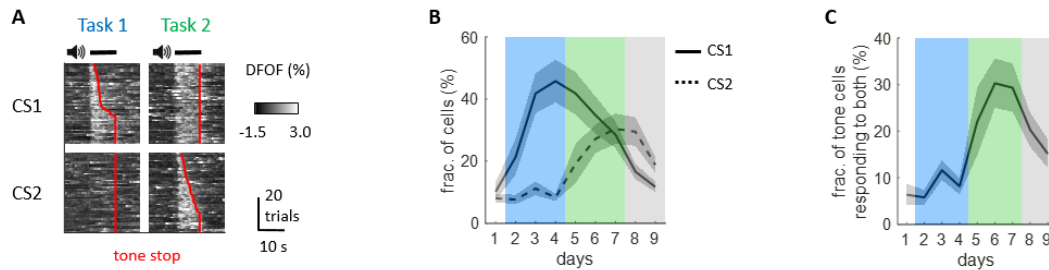


FIGURE 5.4: Quantification of tone-responsive cells. **(A)** Responses to the two tones in the two tasks for an example cell. For CS1 in task 1 and CS2 in task 2, tones could be ended by avoidance shuttles and trials are sorted by tone length. **(B)** Fraction of cells that respond to CS1 (solid line) and CS2 (dashed line) over days. Mean \pm SEM, $n=10$ mice. **(C)** Fraction of tone-responsive cells, that show responses to both tones. Mean \pm SEM, $n=10$ mice.

5.3 Single cell tone responses and decoding of tone-related information

Over the 10 animals that successfully completed the two tasks, we recorded the activity of 2442 neurons (244 ± 62 per subject). While for the 2DAA paradigm the focus was on behavior, here we analyze the responses to the two tones in more detail. We start by considering single cell responses and then use population-level decoding to ask how tones were represented at the time scales of days and milliseconds. Finally, we again consider behavior related signals in the form of predictive information on upcoming avoidance actions.

To investigate how the two tones were represented in mPFC activity over the course of the 2TAA experiment, we first analyzed the activity of individual cells. Fig. 5.4A displays an example cell that responded to CS1, but not CS2 in task 1, but shows responses to both tones in task 2. A quantification of tone selectivities over all neurons showed that mPFC tone responses matched the behavioral discrimination of the two tones: CS1-evoked responses appeared in task 1 and faded away again in task 2, and CS2 responses only became substantial in task 2, where CS2 was first paired with the shock and mice started displaying avoidance responses (Fig. 5.4B). Interestingly, a substantial fraction of cells showed responses to both tones in task 2 (Fig. 5.4C) and we further investigate representational similarity between the two tones in more detail below. As in our other experiments, we next asked how well we could decode

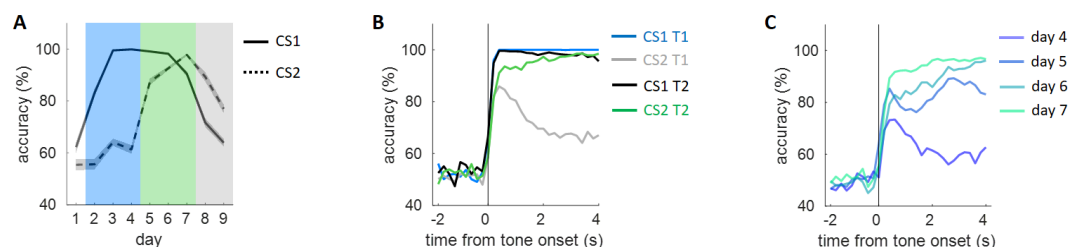


FIGURE 5.5: Tone vs. baseline decoding. **(A)** Accuracy of tone vs. baseline decoders trained per day for CS1 and CS2 presentations. Mean \pm SEM over 20 bootstrap runs. **(B)** Time-resolved tone vs. baseline decoding per task for the two tones. Mean over 20 bootstrap runs. **(C)** Time-resolved tone vs. baseline decoding for CS2 trained on data from days 4 to 6. Mean over 50 bootstrap runs.

tone vs. baseline time steps and found that decoding accuracy closely reflected the fraction of tone-responsive cells over sessions (Fig. 5.5A). This time course again demonstrates the dependence of mPFC tone responses on behavioral relevance and additionally shows that this relevance is evaluated in a tone specific manner.

The analysis of time-resolved tone decoding revealed that CS2-related information had interesting dynamics during task 1 and task 2 tone presentations. In task 1, CS2 decoding accuracy peaked shortly after tone onset and then decayed over the first two seconds (Fig. 5.5B). In task 2, CS2 decoding generally worked better than in task 1, but in contrast to CS1 decoding performance did not reach 100% right after tone start, but was slightly worse and performance increased over the course of the tone presentation. We analyzed this result in more detail by performing CS2 per-time step decoding for individual sessions around the task switch (Fig. 5.5C). Due to the lower number of trials available for training, performance was generally worse, but we found that over task 2 sessions CS2 decoding accuracy was similar at tone onset, but quickly diverged, either decreasing or increasing over the course of the tone presentation. The progressive increase of CS2-related information over days suggests that this effect is based on learning-related changes in the processing of CS2 stimuli.

5.4 Tone identity decoding

Given the overlap of tone responses in task 2 (Fig. 5.4C) we further investigated representational similarity by asking whether responses to the two tones were sufficiently different to allow the decoding of tone identity from neural activity. We first trained tone ID decoders for individual days and observed that tone identity could only be reliably decoded on days 3 and 4, where the discrepancy between CS1-responsive and CS2-responsive cells is maximal (Fig. 5.6A). On days 6 and 7, where tone vs. baseline decoding accuracy was greater than 90% for both tones individually (Fig. 5.5A), tone identity decoding accuracy dropped to below 75%. This discrepancy suggests, that while some tone-specific information exists, the two tones were represented in a similar way. To further investigate this similarity and its temporal evolution, we trained tone ID decoders per time step. We found that the presence of tone-specific information was maximal right after tone onset, and decayed with the duration of the tone (Fig. 5.6B). Interestingly, the time course of this

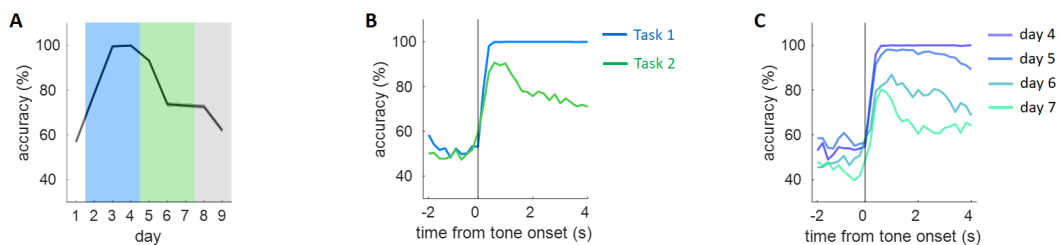


FIGURE 5.6: Tone ID decoding. **(A)** Decoding accuracy of tone identity decoders trained on data from different days. Mean \pm SEM over 50 bootstrap runs. **(B)** Time-resolved decoding of tone identity for SVMs trained on data from tasks 1 and 2. Mean over 50 bootstrap runs. **(C)** Time-resolved decoding of tone identity for days 4 to 6. Mean over 50 bootstrap runs.

decay followed the decay of CS2-related information in task 1 (Fig. 5.5B), suggesting there might be similar information processing dynamics. When repeating the same analysis for individual days, we found that tone identity decoding continually gets worse over task 2 sessions and that the decay of tone identity information becomes faster over days (Fig. 5.6C). Surprisingly, these results show that the representational similarity of the two tones does not follow behavioral discrimination. While animals have similar shuttle levels in response to CS1 and CS2 on day 6, on day 7 there is a clear difference between tone-induced shuttling for the two tones (Fig. 5.2B). This indicates that tone representation is not directly linked to behavior execution, which is consistent with our findings in the 2DAA experiment, where we do not find clear differences in tone-related activity between avoidance and error trials.

5.5 Avoidance action prediction

We next asked if neural activity from the 2TAA experiment contained similar levels of action predictive information as for the 2DAA experiment. Importantly, we can use the comparison between the two experiments as a way to estimate the existence of action-specific signals in the 2DAA experiment. As avoidance actions stay the

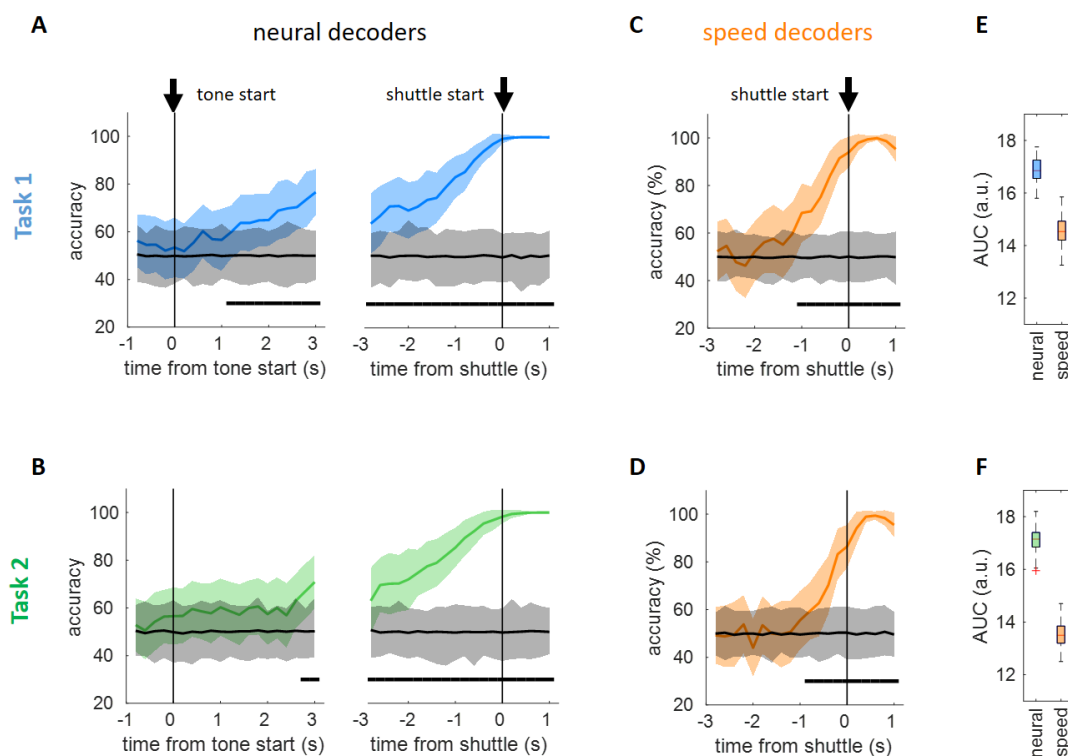


FIGURE 5.7: Prediction of avoidance actions in the 2TAA experiment. **(A)** Time-resolved decoding of trial-type (avoid or error) for time steps aligned to tone start (left) or shuttle action start (right). Decoding performance (blue) is displayed as mean \pm SD over 100 bootstrap runs. Black line and gray area represent mean and 0.5th and 99.5th percentiles of the distribution of mean accuracies of 200 shuffle repetitions of a permutation test (see methods). Black line at the bottom indicates significance ($p < 0.01$). **(B)** Same as (A) for task 2. **(C)** Trial type decoding based on motion information (same as right plot in (A) otherwise). **(D)** Same as (C) for task 2. **(E)** Quantification of decoding performance over 100 bootstrap runs. Area under the curve (AUC) is calculated over the whole period from 3s before to 1s after action start. **(F)** Same as (E) for task 2.

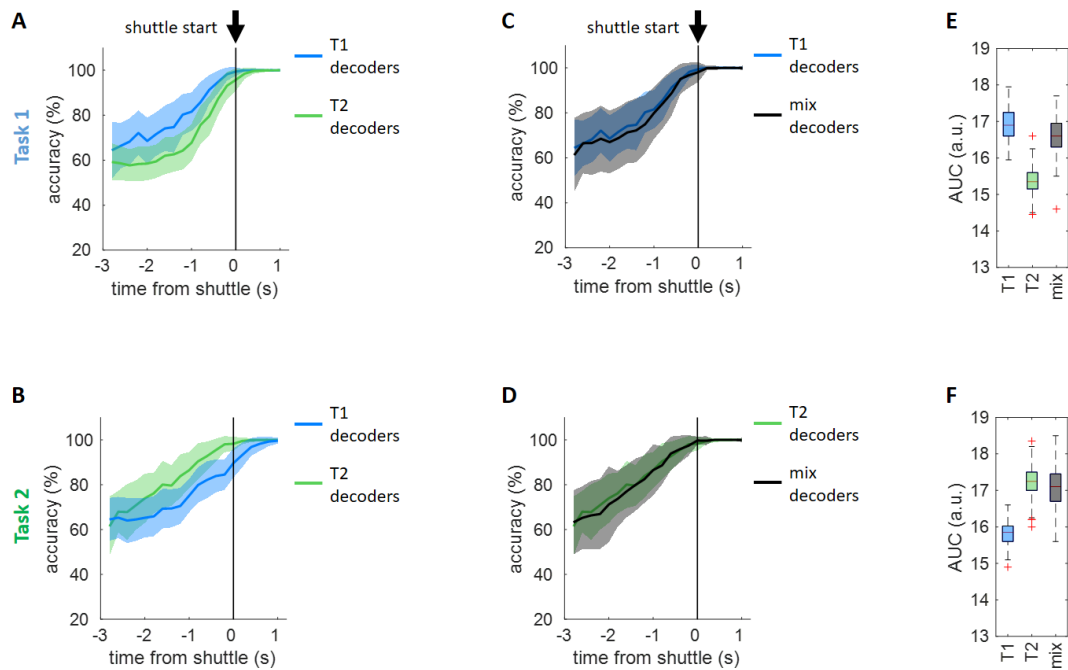


FIGURE 5.8: Action prediction performance evaluated across tasks in the 2TAA experiment. **(A)** Prediction of trial types for task 1 trials, using decoders trained on task 1 (within task testing) or trained on task 2 (across task testing). **(B)** Same as (A) for task 2 trials. **(C)** Comparison of within-task performance with decoders trained using data from both tasks (mix decoders). **(D)** Same as (C) for task 2 trials, mix decoders are the same for (C) and (D). **(E,F)** Quantification of decoder performance for the 3 decoder sets ($n = 100$ bootstrap runs).

same in the 2TAA experiment, any task-related differences that are present in 2DAA, but not 2TAA, are likely to be related to the change of action.

We applied the previously introduced decoding analysis for discriminating avoidance and error trials and found that the results were highly similar between the two experiments. As in the 2DAA experiment, avoidance actions could be decoded above chance levels already 3s before action onset in both tasks (Fig. 5.7A,B). The comparison to speed-based decoding also showed similar accuracies levels that stayed clearly below the performance of neural decoders (Fig. 5.7C-F). This similarity between the results from the two experiments highlights the robustness of the used analysis.

Across-task decoding and mixed decoding using data from both tasks also produced highly similar results to the 2DAA experiment (Fig. 5.8). These results again show that performance differences in across-task decoding (Fig. 5.8A,B) do not originate from conceptual differences between the executed actions in the two tasks. This analysis further strengthens the interpretation that the two actions in 2DAA are associated with similar activity patterns in mPFC.

Chapter 6

Discussion

In this thesis we investigated the role of mPFC in linking sensory stimuli to conditioned behaviors. We approached this question by recording mPFC activity over the course of different conditioning paradigms with changing relationships between stimuli and behavior. In particular, we asked how sensory responses evolve with learning and how they are related to behavior execution.

In all experiments we found clear tone responses that changed over the course of learning. Tone responses emerged with conditioning, were selective to behaviorally relevant tones and rapidly disappeared as soon as stimuli ceased to induce fear-related behaviors. The identification of behavior-related activity was more challenging, as general motion-related activity made it difficult to isolate signals that are related to specific behaviors. However, using a decoding approach, we found that mPFC activity carries predictive information regarding upcoming avoidance actions already several seconds before action-onset. This predictive activity was not specific to different learned avoidance actions, but rather remained similar irrespective of the identity of the executed action. In general, we did not find any evidence of a clear link between tone-related activity and activity related to behavior execution in either of the experimental paradigms.

In the following sections we discuss the results of our three experiments in detail and highlight common themes and findings as well as differences. Finally we discuss limitations of our experimental approach, give an outlook on future work and present conclusions of the work conducted in the scope of this thesis.

6.1 Evolution of prefrontal tone responses over fear conditioning and extinction

We recorded and analyzed the activity of thousands of prefrontal excitatory cells over the full duration of a 6-day fear conditioning paradigm. Our recording and analysis procedures allowed us to track all recorded cells over the course of the six recording sessions. We found that previously unresponsive cells developed a tone response during and after the conditioning session on day 3 (Fig. 3.3), which often remained stable over all post-conditioning days (Figs. 3.3, 3.5). We analyzed the relation of tone responses to freezing behavior and found that there was no correlation between the level of activity of tone-responsive cells and freezing behavior on the second to second basis (Fig. 3.4).

This finding is at odds with interpretations presented in previous work. Burgos-Robles et al. [69] proposed that PL drives freezing via the sustained activity of tone-responsive cells. The authors base this interpretation on a temporal correlation of tone responses and freezing behavior. However, their observation that mPFC tone responses correlate with freezing is not surprising, as tone presentations and freezing behaviors have a systematic temporal overlap inherent to task design (see Fig. 3.2B). This temporal overlap generates a high baseline correlation between tone responses and freezing and a link between tone-evoked activity and freezing would only be implied by a further correlation outside of tone presentations, which seems limited in their data. We quantified potential joint coding of tone information and freezing by testing for correlations between tone z-scores (limited to tone onset to address the overlap issue) and freezing scores, and found that there was no substantial correlation between how cells responded to the tone and related to freezing (Fig. 3.4).

Yet, consistent with Burgos-Robles et al. [69], we found that tone responses disappeared as animals stopped freezing during extinction sessions (Fig. 3.6). Our decoding analysis showed that individual trials in which animals did not freeze contained less tone-related information than freezing trials. This trial-to-trial correlation suggests that there is some link between prefrontal tone responses and behavior execution, but we did not find a specific neural representation of this link. Thus it remains unclear whether this link is causal and what mechanisms underlie the transformation of tone-related activity into signals driving freezing behavior.

In a series of papers, Courtin et al. investigated this mechanistic basis using electrode recordings and optogenetic manipulations [73, 74, 75]. They found that a 4Hz rhythm in mPFC was causally related to the expression of freezing behavior. Although the authors could demonstrate the importance of the 4Hz rhythm using temporally precise manipulations of neural activity, it remained unclear what the origin of this rhythm is and how it relates to CS presentations. More recently, a study found that the 4Hz rhythm originates from regular breathing during freezing and is transmitted to mPFC via the olfactory bulb [114]. Based on this finding, Bagur et al. [115] showed that prefrontal activity is predominantly important for freezing maintenance rather than freezing initiation. The initiation of freezing is crucially dependent on amygdala circuits, but it remains unclear what the relevance of CS+ evoked activity in mPFC is for the initiation of freezing. Unfortunately, due to the low temporal resolution of calcium imaging, such rhythms were not visible in our data.

Overall, although our findings are consistent with prefrontal CS+ responses driving the execution of freezing behavior, we did not find any signatures of mechanisms linking tone-related activity to freezing. However, another possible explanation would be that prefrontal tone responses are tightly coupled to behavioral relevance but do not directly cause behavior. This would also explain the lack of tone responses in trials without freezing, as the absence of freezing indicates a loss of behavioral relevance.

6.2 Neural correlates of tones and actions in active avoidance

We designed two complementary active avoidance paradigms in which we adapted the conditioned response and the conditioned tone respectively. We used these adapted mappings from tone to behavior to assess concurrent changes in mPFC activity. Here we first discuss general properties of the neural activity related to active avoidance paradigms and then consider the effects of the task switches in the 2-dimensional active avoidance (2DAA) and 2-tone active avoidance (2TAA) experiments.

Similar to the fear conditioning paradigm, tone responses were easy to identify due to their clear temporal alignment to tone on- and offset (Fig. 4.7). Compared to fear conditioning, a larger fraction of cells was classified as tone-responsive. This can however be explained by the higher number of trials and the resulting increase in statistical power of tone cell classification procedure. Moreover, tone responses stayed stable over multiple sessions, but disappeared in extinction sessions.

Identifying activity patterns related to active avoidance is complicated by neural correlates of general motion [116]. Consistent with recent studies demonstrating widespread activity related to movements [117, 118, 119], we found that the activity of many prefrontal cells was modulated by motion (Fig. 3.4). We approached this challenge by focusing on pre-action activity in order to investigate the cognitive process leading to the execution of avoidance actions. We aligned trials to the onset of avoidance actions (Fig. 4.5) and found that there were only subtle differences between avoidance trials and error trials in individual cells (Fig. 4.9). This contrasts our results in the fear conditioning paradigm, where we found that differences in freezing behavior between trials were reflected by responses to the tone. However, as stated above, we interpret this correlation between tone responses and freezing behavior as a reflection of the temporal evolution of the tone's behavioral relevance. Within active avoidance sessions, however, the behavioral relevance of the tone remains constantly high, and failures to avoid are not based on the failure to recognize the behavioral relevance of the tone. Especially after initial learning, error trials are rather a reflection of the conflict between freezing and avoidance actions.

Despite the subtlety of avoidance-predictive activity in single cells, we found that population activity contained predictive information about avoidance actions up to 3s before action onset by following the decoding approach from Jercog et al. [99] (Fig. 4.10). Importantly, this predictive information cannot be explained by motion correlates, as decoders trained on motion information from video tracking only achieve above-chance decoding accuracies 2s after neural decoders. While the results in the study of Jercog et al. [99] are qualitatively similar, our data show substantially higher effect sizes in the comparison between neural and speed decoders (3s vs. 1s instead of 750ms vs. 500ms). This increased performance of neural decoders is probably based on two factors. First, we recorded from a substantially higher number of cells (ca. 3000 vs. ca. 600). Second, since we recorded from the same neurons over different sessions, we could pool trials from different sessions, increasing the number of data points that can be used for training SVM decoders.

Using this decoding approach, Jercog et al. [99] showed that cells which have an excitatory response to tone presentations are better predictors of avoidance actions than other cells. The authors argue that this result provides a link between tone

responses and the execution of avoidance actions. However, the relevance of this difference between tone-excited and other cells found in their study is questionable, as the predictive power of tone-excited cells did not exceed the predictive power of motion-related information. Indeed, we did not find an effect that links tone responses to avoidance-predictive activity in our data. Our removal analysis shows that even when removing 90% of the most tone-responsive cells, good avoidance action prediction can still be achieved (Fig. 4.12). By contrast, when removing cells with a high avoidance score, a large drop in accuracy can be observed, indicating that avoidance-related information is not jointly coded with tone information.

Based on our results, it does not seem appropriate to attribute the transformation from sensory information to action initiation to local processing in mPFC, which is often the narrative in current literature. However, it is clear that mPFC plays an important role in the distributed network that is involved in this transformation. In particular, mPFC projections to the BLA and ventral striatum have been shown to differentially modulate avoidance behavior [100]. Nevertheless, projection-specific imaging of mPFC cells projecting to BLA and striatum did not find differences in tone responses in these subpopulations [120], further highlighting the difficulty of linking sensory-evoked responses to action initiation. Ultimately, further work is required to investigate how the coordination between different brain areas leads to action initiation, e.g. through the study of cortico-striatal loops [121]. Such a distributed transformation of sensory-driven activity to drive avoidance behavior would be consistent with the slow rise of avoidance-predictive information we observed in our decoding analysis (Fig. 4.10A).

6.3 Lack of action-specificity in the 2DAA paradigm

The 2DAA paradigm was based on the idea that if the action induced by the tone changes, we might see a change in prefrontal activity that gives insights into how tone and action are linked. However, our analyses did not show a change in activity that could be clearly traced back to the switch between the two avoidance actions. While there were some differences between tasks, it was generally difficult to assign these differences to the task switch.

One challenge we faced was that the two conditioned responses were present in different phases of the experiment. Thus, when analyzing action specificity, one has to control that the observed changes did not simply occur due to temporal drifts. One strategy to do that is to see how task-wise changes relate to session-wise changes. If the change at the task switch is bigger than the change between sessions, one could conclude that there are task-specific changes. We followed this strategy when comparing tone responses between the two tasks (Figs. 4.7, 7.8) and found a small effect when comparing tone z-scores of different sessions. However, the effect might still be explained by a general stabilization of tone responses over time. Furthermore, the potential of this approach is limited by the fact that any present effect is probably diluted, because changes that are associated with learning in task 2 do not occur abruptly but rather develop slowly and at different speeds for different animals.

Another approach we followed was to use the two avoidance experiments (2DAA and 2TAA) as mutual comparisons, as the manipulations between tasks were different. One question that we addressed this way was whether the differences between

within-task and across-task decoding of avoidance actions were related to the difference between actions in the 2DAA experiment (Fig. 4.11). When performing the same analysis for the 2TAA experiment, we found that the results were almost identical (Fig. 5.8). Since the action taken in the two tasks is the same for the 2TAA experiment, we concluded that the differences we found in the 2DAA experiment were not based on action-specificity.

It is important to note that the fact that we did not find action specificity in our experiments does not mean that such specificity does not exist in mPFC. Next to the methodological difficulties of identifying such specificity described above, one possible explanation for the absence of action-specific neural activity could be that the two avoidance actions were too similar. While animals did adapt their actions (Fig. 4.2), they might not have learned that changing the direction of their shuttle motion corresponds to a conceptually different action. It would be interesting to see if we could find differences in action-predictive activity if avoidance was based on two very different actions (e.g. shuttling vs. lever pressing). However, based on our experience in training animals on avoidance tasks, it might be difficult to find a setting in which animals can learn such a task.

In general, investigating cognitive flexibility in the context of aversive learning seems especially challenging as flexibility is systematically reduced when animals are in threatening situations [122]. When running behavioral pilot studies during the development of the 2DAA experiment, we experimented with various versions of tasks combining multiple tones and multiple actions. However, we experienced difficulties in training mice in such paradigms. Additionally, as discussed above, it would have been advantageous to switch back and forth between different tasks, as revisiting tasks allows to control for temporal effects [88]. Yet the conceptually relatively simple 2DAA and 2TAA experiments already pushed the limits of what mice could learn and the long duration of the experiments introduced challenges for the steady recording of neural activity, which precluded prolonging tasks even further.

6.4 Representational similarity of avoidance-inducing tones

In the 2TAA experiment we trained mice to shuttle in response to two different tones in two consecutive tasks. We were particularly interested in the evolution of tone responses upon the task-switch. Consistent with previous experiments, CS2, which was first paired with the shock in the second task, did not evoke strong responses in task 1 (Fig. 5.4). In task 2 both tones elicited clear responses, although CS1 responses declined, while CS2 responses increased throughout the task. We investigated the similarity of these tone responses and found that it increased over days, as reflected by the decreasing accuracy values in the tone identity decoding analysis (Fig. 5.6).

We further wanted to understand whether this representational similarity was based on the two tones being linked to the same conditioned behavior. One conceptual issue with this question was that we could not evaluate this similarity in comparison to a baseline setting. In other words, it is not clear what level of similarity one should expect from two tones that are linked to different behaviors. However, our time-resolved decoding analysis of tone identity offered a solution to this problem, as it allowed us to compare similarities of tone responses at different time steps. Using this analysis, we found that, over the course of tone presentations, similarity increased (Fig. 5.6).

This finding suggests that the activity evoked by the two tones converges to a joint threat representation. Consistent with this idea, CS1 continued to induce shuttling throughout task 2, after it had lost its connection to the shock. In contrast, CS2-induced shuttling decays much more rapidly upon the transition to extinction session (Fig. 5.2). This suggests that the threat association to CS1 is maintained in task 2 because of the ongoing danger of receiving a shock in CS2 trials. A further interesting finding is that CS1 and CS2 responses became more similar even as behavioral discrimination between the tones increased at the end of task 2 (Fig. 5.6C). This is similar to the finding that tone responses did not clearly differ between avoidance and error trials and further suggests that there is no direct link between tone-evoked activity and behavior execution.

An interesting difference between the 2DAA and 2TAA experiments was that for 2TAA tone responses were generally weaker in task 2 (Fig. 5.4), whereas they increased in intensity for 2DAA (Fig. 4.7). This observation could be related to differences between the type of learning required in the two experiments. While the 2DAA experiment requires the adaptation of a previously learned action, 2TAA requires transferring a learned action to another sensory stimulus, which might require different learning mechanisms.

Finally, another difference between the two experimental paradigms was the rate of animals successfully learning both tasks. In the 2TAA experiment all animals learned the first task, but 6 out of the 16 animals did not manage to learn the second task. In contrast, all 2DAA animals learned both tasks successfully. However, within the 2TAA experiment, the success rate varied between different recording groups. We usually worked with groups of animals that were co-housed in one cage for every group of animals that entered the paradigm (maximally 5 at a time). Most of the unsuccessful animals came from two groups that were born at the same time and recorded from in close succession. It is thus likely that some factors that were specific to these two cages caused the lower success rate. Potential sources of lower performance could be stressors in the housing conditions such as noise, changes in temperature or humidity, or issues with breeding.

6.5 Limitations and outlook

In this thesis, we investigated the neural basis of stimulus-response mapping by manipulating the mappings that animals learn through conditioning, while recording neural activity. The usefulness of this approach requires that the manipulations of the link between stimuli and behavior are reflected in the recorded neural activity. This was the case for the 2TAA experiment and the change between avoidance-inducing tones allowed us to assess how the two tone representations were related. In the 2DAA experiment however, the lack of action-specific activity precluded investigations of the link between the tone and different actions. Thus, the approach of manipulation stimulus-response mappings was only partially successful.

The use of calcium imaging allowed us to record the activity of a substantially higher number of neurons compared to previous work using electrophysiological approaches. However, this advantage comes at the cost of a lower temporal resolution. Especially in the fear conditioning experiments, it would have been interesting to further investigate oscillations during freezing behavior and their relation to tone-evoked activity.

As introduced in section 1.1.5, there are many ways to analyze neural data and, over the course of this thesis, we have explored various approaches. The presented decoding approach was particularly useful because it allowed us to ask clear questions regarding the representation of experimental variables and how they change over time. However, there are also clear limitations to analyses based on decoding. For example, while one can ask if a decoder can find differences between two data sets, it is not always straightforward to understand what these differences are based on. An example of this issue is the distinction of tone and baseline time steps for avoidance trials. Towards the end of avoidance trials, there are two factors that discriminate trials from baseline (tone and behavior) and these factors are difficult to disentangle using decoding. In such settings, model-based approaches that isolate the relation of neural activity to individual signals of interest can be helpful. However, when many signals overlap, it can be difficult to properly fit models such that they appropriately disentangle different contributions to changes in neural activity. In our case, model fitting was difficult as there were many systematic overlaps including tone presentations, avoidance actions, escape actions, shocks presentations and general motion (Fig. 4.6). Due to these factors, it was difficult to leverage the power of model-based approaches and we did not include such analyses in this thesis, as they did not provide clear answers to the questions addressed here.

Finally, another limitation of the presented study, but also of the field in general, is a lack of theoretical understanding of what links between sensory stimuli and action-related signals could look like. The narrative of various studies is that there is some form of joint coding of sensory information and action-inducing activity in individual cells [69, 78, 99]. However, such joint coding is only one simple model of how sensory and behavioral signals can interact in the studied networks of neurons. Since we did not find evidence for joint coding, we believe that further theoretical investigations are necessary to generate predictions that can be tested using experimental data. Importantly, such predictions can motivate and guide the use of manipulation techniques such as optogenetics and increase the relevance of the displayed effects on behavior.

6.6 Conclusion

Throughout our three experimental paradigms, we observed that prefrontal tone responses are tightly coupled to behavioral relevance. However, we did not find a link between tone-evoked responses and the neural correlates of fear-related behaviors. Nevertheless, we found that mPFC contained information that could be used to predict avoidance actions up to three seconds before action initiation. Yet, it remained unclear how this predictive information is linked to tone-responsive cells. Taken together, our results suggest that stimuli and conditioned responses are not linked through a simple joint coding mechanism in mPFC and motivate theoretical investigations of how sensory stimuli are transformed into action initiation signals over networks of interconnected brain regions.

Bibliography

- [1] J. E. LeDoux. "Emotion circuits in the brain". In: *Annual Review of Neuroscience* 23 (2000), pp. 155–184. DOI: [10.1146/annurev.neuro.23.1.155](https://doi.org/10.1146/annurev.neuro.23.1.155).
- [2] Patricia H. Janak et al. "From circuits to behaviour in the amygdala". In: *Nature* 517.7534 (Jan. 2015), pp. 284–292. DOI: [10.1038/nature14188](https://doi.org/10.1038/nature14188).
- [3] I. P. Pavlov. *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. Oxford Univ. Press. Oxford, England, 1927.
- [4] Charles B. Ferster et al. "Schedules of reinforcement." In: *Appleton-Century-Crofts* (1957). DOI: <https://doi.org/10.1037/10627-000>.
- [5] RA Rescorla et al. "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement". In: *Classical Conditioning II: Current Research and Theory*. Vol. Vol. 2. Jan. 1972.
- [6] A. L. Hodgkin et al. "A quantitative description of membrane current and its application to conduction and excitation in nerve". In: *The Journal of Physiology* 117.4 (1952), pp. 500–544. DOI: [10.1113/jphysiol.1952.sp004764](https://doi.org/10.1113/jphysiol.1952.sp004764).
- [7] Eric R. Kandel. "The Molecular Biology of Memory Storage: A Dialogue Between Genes and Synapses". In: *Science* 294.5544 (Nov. 2001), pp. 1030–1038. DOI: [10.1126/science.1067020](https://doi.org/10.1126/science.1067020).
- [8] D. H. Hubel et al. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". In: *The Journal of Physiology* 160.1 (Jan. 1962), pp. 106–154.2.
- [9] Marie Carlén. "What constitutes the prefrontal cortex?" In: *Science* 358.6362 (Oct. 2017), pp. 478–482. DOI: [10.1126/science.aan8868](https://doi.org/10.1126/science.aan8868).
- [10] Dirk Feldmeyer et al. "Barrel cortex function". In: *Progress in Neurobiology*. Conversion of Sensory Signals into Perceptions, Memories and Decisions 103 (Apr. 2013), pp. 3–27. DOI: [10.1016/j.pneurobio.2012.11.002](https://doi.org/10.1016/j.pneurobio.2012.11.002).
- [11] Andrew J. Peters et al. "Emergence of reproducible spatiotemporal activity during motor learning". In: *Nature* 510.7504 (June 2014), pp. 263–267. DOI: [10.1038/nature13235](https://doi.org/10.1038/nature13235).
- [12] Edvard I. Moser et al. "Place Cells, Grid Cells, and the Brain's Spatial Representation System". In: *Annual Review of Neuroscience* 31.1 (2008), pp. 69–89. DOI: [10.1146/annurev.neuro.31.061307.090723](https://doi.org/10.1146/annurev.neuro.31.061307.090723).
- [13] Anthony Dickinson et al. "Motivational control of goal-directed action". In: *Animal Learning & Behavior* 22.1 (1994), pp. 1–18.
- [14] Robert J. Blanchard et al. "Crouching as an index of fear". In: *Journal of Comparative and Physiological Psychology* 67.3 (1969), pp. 370–375. DOI: [10.1037/h0026779](https://doi.org/10.1037/h0026779).

- [15] J. E. LeDoux et al. "The birth, death and resurrection of avoidance: a reconceptualization of a troubled paradigm". In: *Molecular Psychiatry* 22.1 (Jan. 2017), pp. 24–36. DOI: [10.1038/mp.2016.166](https://doi.org/10.1038/mp.2016.166).
- [16] O. H. Mowrer. "Two-factor learning theory: summary and comment". In: *Psychological Review* 58.5 (1951), pp. 350–354. DOI: [10.1037/h0058956](https://doi.org/10.1037/h0058956).
- [17] Angelos-Miltiadis Kryptos et al. "Avoidance learning: a review of theoretical models and recent developments". In: *Frontiers in Behavioral Neuroscience* 9 (July 2015). DOI: [10.3389/fnbeh.2015.00189](https://doi.org/10.3389/fnbeh.2015.00189).
- [18] Robert C. Bolles. "Species-specific defense reactions and avoidance learning". In: *Psychological Review* 77.1 (1970), pp. 32–48. DOI: [10.1037/h0028589](https://doi.org/10.1037/h0028589).
- [19] Martin E. Seligman et al. "A cognitive theory of avoidance learning". In: *Contemporary approaches to conditioning and learning*. Oxford, England, 1973, pp. xii, 321–xii, 321.
- [20] Peter Lovibond. "Fear and Avoidance: An Integrated Expectancy Model". In: *Fear and learning: From basic processes to clinical implications*. Washington, DC, US, 2006, pp. 117–132. DOI: [10.1037/11474-006](https://doi.org/10.1037/11474-006).
- [21] Robert A. Rescorla et al. "Inhibition of avoidance behavior". In: *Journal of Comparative and Physiological Psychology* 59.3 (1965), pp. 406–412. DOI: [10.1037/h0022060](https://doi.org/10.1037/h0022060).
- [22] Christopher K. Cain. "Avoidance Problems Reconsidered". In: *Current Opinion in Behavioral Sciences* 26 (Apr. 2019), pp. 9–17. DOI: [10.1016/j.cobeha.2018.09.002](https://doi.org/10.1016/j.cobeha.2018.09.002).
- [23] Karl Deisseroth. "Optogenetics: 10 years of microbial opsins in neuroscience". In: *Nature Neuroscience* 18.9 (Sept. 2015), pp. 1213–1225. DOI: [10.1038/nn.4091](https://doi.org/10.1038/nn.4091).
- [24] Bryan L. Roth. "DREADDs for Neuroscientists". In: *Neuron* 89.4 (Feb. 2016), pp. 683–694. DOI: [10.1016/j.neuron.2016.01.040](https://doi.org/10.1016/j.neuron.2016.01.040).
- [25] György Buzsáki. "Large-scale recording of neuronal ensembles". In: *Nature Neuroscience* 7.5 (May 2004), pp. 446–451. DOI: [10.1038/nn1233](https://doi.org/10.1038/nn1233).
- [26] Elizabeth J. O. Hamel et al. "Cellular Level Brain Imaging in Behaving Mammals: An Engineering Approach". In: *Neuron* 86.1 (Apr. 2015), pp. 140–159. DOI: [10.1016/j.neuron.2015.03.055](https://doi.org/10.1016/j.neuron.2015.03.055).
- [27] James J. Jun et al. "Fully integrated silicon probes for high-density recording of neural activity". In: *Nature* 551.7679 (Nov. 2017), pp. 232–236. DOI: [10.1038/nature24636](https://doi.org/10.1038/nature24636).
- [28] Christine Grienberger et al. "Imaging Calcium in Neurons". In: *Neuron* 73.5 (Mar. 2012), pp. 862–885. DOI: [10.1016/j.neuron.2012.02.011](https://doi.org/10.1016/j.neuron.2012.02.011).
- [29] Tobias Rose et al. "Putting a finishing touch on GECIs". In: *Frontiers in Molecular Neuroscience* 7 (2014).
- [30] Anthony Holtmaat et al. "Long-term, high-resolution imaging in the mouse neocortex through a chronic cranial window". In: *Nature Protocols* 4.8 (Aug. 2009), pp. 1128–1144. DOI: [10.1038/nprot.2009.89](https://doi.org/10.1038/nprot.2009.89).
- [31] Fritjof Helmchen et al. "Deep tissue two-photon microscopy". In: *Nature methods* 2.12 (2005), pp. 932–940.

- [32] Michael J. Levene et al. "In Vivo Multiphoton Microscopy of Deep Brain Tissue". In: *Journal of Neurophysiology* 91.4 (Apr. 2004), pp. 1908–1912. DOI: [10.1152/jn.01007.2003](https://doi.org/10.1152/jn.01007.2003).
- [33] Daniel A. Dombeck et al. "Imaging Large-Scale Neural Activity with Cellular Resolution in Awake, Mobile Mice". In: *Neuron* 56.1 (Oct. 2007), pp. 43–57. DOI: [10.1016/j.neuron.2007.08.003](https://doi.org/10.1016/j.neuron.2007.08.003).
- [34] Christopher D. Harvey et al. "Intracellular dynamics of hippocampal place cells during virtual navigation". In: *Nature* 461.7266 (Oct. 2009), pp. 941–946. DOI: [10.1038/nature08499](https://doi.org/10.1038/nature08499).
- [35] Georg B. Keller et al. "Sensorimotor Mismatch Signals in Primary Visual Cortex of the Behaving Mouse". In: *Neuron* 74.5 (June 2012), pp. 809–815. DOI: [10.1016/j.neuron.2012.03.040](https://doi.org/10.1016/j.neuron.2012.03.040).
- [36] Fritjof Helmchen et al. "A Miniature Head-Mounted Two-Photon Microscope: High-Resolution Brain Imaging in Freely Moving Animals". In: *Neuron* 31.6 (Sept. 2001), pp. 903–912. DOI: [10.1016/S0896-6273\(01\)00421-4](https://doi.org/10.1016/S0896-6273(01)00421-4).
- [37] Kunal K. Ghosh et al. "Miniaturized integration of a fluorescence microscope". In: *Nature Methods* 8.10 (Oct. 2011), pp. 871–878. DOI: [10.1038/nmeth.1694](https://doi.org/10.1038/nmeth.1694).
- [38] Yaniv Ziv et al. "Long-term dynamics of CA1 hippocampal place codes". In: *Nature Neuroscience* 16.3 (Mar. 2013), pp. 264–266. DOI: [10.1038/nn.3329](https://doi.org/10.1038/nn.3329).
- [39] Benjamin F. Grewe et al. "Neural ensemble dynamics underlying a long-term associative memory". In: *Nature* 543.7647 (Mar. 2017), pp. 670–675. DOI: [10.1038/nature21682](https://doi.org/10.1038/nature21682).
- [40] L Paninski et al. "Neural data science: accelerating the experiment-analysis-theory cycle in large-scale neuroscience". In: *Current Opinion in Neurobiology*. Neurotechnologies 50 (June 2018), pp. 232–241. DOI: [10.1016/j.conb.2018.04.007](https://doi.org/10.1016/j.conb.2018.04.007).
- [41] Mattia Rigotti et al. "The importance of mixed selectivity in complex cognitive tasks". In: *Nature* 497.7451 (May 2013), pp. 585–590. DOI: [10.1038/nature12160](https://doi.org/10.1038/nature12160).
- [42] Jonathan W. Pillow et al. "Spatio-temporal correlations and visual signalling in a complete neuronal population". In: *Nature* 454.7207 (Aug. 2008), pp. 995–999. DOI: [10.1038/nature07140](https://doi.org/10.1038/nature07140).
- [43] Il Memming Park et al. "Encoding and decoding in parietal cortex during sensorimotor decision-making". In: *Nature Neuroscience* 17.10 (Oct. 2014), pp. 1395–1403. DOI: [10.1038/nn.3800](https://doi.org/10.1038/nn.3800).
- [44] Caroline A. Runyan et al. "Distinct timescales of population coding across cortex". In: *Nature* 548.7665 (Aug. 2017), pp. 92–96. DOI: [10.1038/nature23020](https://doi.org/10.1038/nature23020).
- [45] Ethan M. Meyers et al. "Dynamic Population Coding of Category Information in Inferior Temporal and Prefrontal Cortex". In: *Journal of Neurophysiology* 100.3 (Sept. 2008), pp. 1407–1419. DOI: [10.1152/jn.90248.2008](https://doi.org/10.1152/jn.90248.2008).
- [46] Mark G. Stokes et al. "Dynamic Coding for Cognitive Control in Prefrontal Cortex". In: *Neuron* 78.2 (Apr. 2013), pp. 364–375. DOI: [10.1016/j.neuron.2013.01.039](https://doi.org/10.1016/j.neuron.2013.01.039).
- [47] Ofer Mazor et al. "Transient Dynamics versus Fixed Points in Odor Representations by Locust Antennal Lobe Projection Neurons". In: *Neuron* 48.4 (Nov. 2005), pp. 661–673. DOI: [10.1016/j.neuron.2005.09.032](https://doi.org/10.1016/j.neuron.2005.09.032).

- [48] Byron M Yu et al. "Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity". In: *Advances in Neural Information Processing Systems*. Vol. 21. 2008.
- [49] John P. Cunningham et al. "Dimensionality reduction for large-scale neural recordings". In: *Nature Neuroscience* 17.11 (Nov. 2014), pp. 1500–1509. DOI: [10.1038/nn.3776](https://doi.org/10.1038/nn.3776).
- [50] Mark M. Churchland et al. "Neural population dynamics during reaching". In: *Nature* 487.7405 (July 2012), pp. 51–56. DOI: [10.1038/nature11129](https://doi.org/10.1038/nature11129).
- [51] Valerio Mante et al. "Context-dependent computation by recurrent dynamics in prefrontal cortex". In: *Nature* 503.7474 (Nov. 2013), pp. 78–84. DOI: [10.1038/nature12742](https://doi.org/10.1038/nature12742).
- [52] Peiran Gao et al. "On simplicity and complexity in the brave new world of large-scale neuroscience". In: *Current Opinion in Neurobiology*. Large-Scale Recording Technology (32) 32 (June 2015), pp. 148–155. DOI: [10.1016/j.conb.2015.04.003](https://doi.org/10.1016/j.conb.2015.04.003).
- [53] Dmitry Kobak et al. "Demixed principal component analysis of neural population data". In: *eLife* 5 (Apr. 2016), e10989. DOI: [10.7554/eLife.10989](https://doi.org/10.7554/eLife.10989).
- [54] Yuanjun Gao et al. "Linear dynamical neural population models through nonlinear embeddings". In: *arXiv:1605.08454 [q-bio, stat]* (Oct. 2016).
- [55] Scott Linderman et al. "Bayesian Learning and Inference in Recurrent Switching Linear Dynamical Systems". In: *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*. Apr. 2017, pp. 914–922.
- [56] Chethan Pandarinath et al. "Inferring single-trial neural population dynamics using sequential auto-encoders". In: *Nature Methods* 15.10 (Oct. 2018), pp. 805–815. DOI: [10.1038/s41592-018-0109-9](https://doi.org/10.1038/s41592-018-0109-9).
- [57] Sevil Duvarci et al. "Amygdala microcircuits controlling learned fear". In: *Neuron* 82.5 (June 2014), pp. 966–980. DOI: [10.1016/j.neuron.2014.04.042](https://doi.org/10.1016/j.neuron.2014.04.042).
- [58] Gregory J. Quirk et al. "Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: Parallel recordings in the freely behaving rat". In: *Neuron* 15.5 (Nov. 1995), pp. 1029–1039. DOI: [10.1016/0896-6273\(95\)90092-6](https://doi.org/10.1016/0896-6273(95)90092-6).
- [59] J. J. Kim et al. "Effects of amygdala, hippocampus, and periaqueductal gray lesions on short- and long-term contextual fear". In: *Behavioral Neuroscience* 107.6 (Dec. 1993), pp. 1093–1098. DOI: [10.1037//0735-7044.107.6.1093](https://doi.org/10.1037//0735-7044.107.6.1093).
- [60] Stephane Ciocchi et al. "Encoding of conditioned fear in central amygdala inhibitory circuits". In: *Nature* 468.7321 (Nov. 2010), pp. 277–282. DOI: [10.1038/nature09559](https://doi.org/10.1038/nature09559).
- [61] Cyril Herry et al. "Switching on and off fear by distinct neuronal circuits". In: *Nature* 454.7204 (July 2008), pp. 600–606. DOI: [10.1038/nature07166](https://doi.org/10.1038/nature07166).
- [62] Cyril Herry et al. "Encoding of fear learning and memory in distributed neuronal circuits". In: *Nature Neuroscience* 17.12 (Dec. 2014), pp. 1644–1654. DOI: [10.1038/nn.3869](https://doi.org/10.1038/nn.3869).
- [63] Francisco Sotres-Bayon et al. "Prefrontal control of fear: more than just extinction". In: *Current Opinion in Neurobiology*. Cognitive neuroscience 20.2 (Apr. 2010), pp. 231–235. DOI: [10.1016/j.conb.2010.02.005](https://doi.org/10.1016/j.conb.2010.02.005).

- [64] Roger Marek et al. "The amygdala and medial prefrontal cortex: partners in the fear circuit". In: *The Journal of Physiology* 591.10 (May 2013), pp. 2381–2391. DOI: [10.1113/jphysiol.2012.248575](https://doi.org/10.1113/jphysiol.2012.248575).
- [65] Thomas F. Giustino et al. "The Role of the Medial Prefrontal Cortex in the Conditioning and Extinction of Fear". In: *Frontiers in Behavioral Neuroscience* 9 (Nov. 2015). DOI: [10.3389/fnbeh.2015.00298](https://doi.org/10.3389/fnbeh.2015.00298).
- [66] Ivan Vidal-Gonzalez et al. "Microstimulation reveals opposing influences of prelimbic and infralimbic cortex on the expression of conditioned fear". In: *Learning & Memory* 13.6 (Nov. 2006), pp. 728–733. DOI: [10.1101/lm.306106](https://doi.org/10.1101/lm.306106).
- [67] Kevin A. Corcoran et al. "Activity in Prelimbic Cortex Is Necessary for the Expression of Learned, But Not Innate, Fears". In: *Journal of Neuroscience* 27.4 (Jan. 2007), pp. 840–844. DOI: [10.1523/JNEUROSCI.5327-06.2007](https://doi.org/10.1523/JNEUROSCI.5327-06.2007).
- [68] Fabricio H. Do-Monte et al. "A temporal shift in the circuits mediating retrieval of fear memory". In: *Nature* 519.7544 (Mar. 2015), pp. 460–463. DOI: [10.1038/nature14030](https://doi.org/10.1038/nature14030).
- [69] Anthony Burgos-Robles et al. "Sustained Conditioned Responses in Prelimbic Prefrontal Neurons Are Correlated with Fear Expression and Extinction Failure". In: *Journal of Neuroscience* 29.26 (July 2009), pp. 8474–8482. DOI: [10.1523/JNEUROSCI.0378-09.2009](https://doi.org/10.1523/JNEUROSCI.0378-09.2009).
- [70] Verena Senn et al. "Long-Range Connectivity Defines Behavioral Specificity of Amygdala Neurons". In: *Neuron* 81.2 (Jan. 2014), pp. 428–437. DOI: [10.1016/j.neuron.2013.11.006](https://doi.org/10.1016/j.neuron.2013.11.006).
- [71] Ekaterina Likhtik et al. "Prefrontal entrainment of amygdala activity signals safety in learned fear and innate anxiety". In: *Nature Neuroscience* 17.1 (Jan. 2014), pp. 106–113. DOI: [10.1038/nn.3582](https://doi.org/10.1038/nn.3582).
- [72] Anthony Burgos-Robles et al. "Amygdala inputs to prefrontal cortex guide behavior amid conflicting cues of reward and punishment". In: *Nature Neuroscience* 20.6 (June 2017), pp. 824–835. DOI: [10.1038/nn.4553](https://doi.org/10.1038/nn.4553).
- [73] Julien Courtin et al. "Prefrontal parvalbumin interneurons shape neuronal activity to drive fear expression". In: *Nature* 505.7481 (Jan. 2014), pp. 92–96. DOI: [10.1038/nature12755](https://doi.org/10.1038/nature12755).
- [74] Nikolaos Karalis et al. "4-Hz oscillations synchronize prefrontal–amygdala circuits during fear behavior". In: *Nature Neuroscience* 19.4 (Apr. 2016), pp. 605–612. DOI: [10.1038/nn.4251](https://doi.org/10.1038/nn.4251).
- [75] Cyril Dejean et al. "Prefrontal neuronal assemblies temporally control fear behaviour". In: *Nature* 535.7612 (July 2016), pp. 420–424. DOI: [10.1038/nature18630](https://doi.org/10.1038/nature18630).
- [76] Hiroshi Makino et al. "Circuit Mechanisms of Sensorimotor Learning". In: *Neuron* 92.4 (Nov. 2016), pp. 705–721. DOI: [10.1016/j.neuron.2016.10.029](https://doi.org/10.1016/j.neuron.2016.10.029).
- [77] Pierre Le Merre et al. "The mouse prefrontal cortex: Unity in diversity". In: *Neuron* 109.12 (June 2021), pp. 1925–1944. DOI: [10.1016/j.neuron.2021.03.035](https://doi.org/10.1016/j.neuron.2021.03.035).
- [78] James M. Otis et al. "Prefrontal cortex output circuits guide reward seeking through divergent cue encoding". In: *Nature* 543.7643 (Mar. 2017), pp. 103–107. DOI: [10.1038/nature21376](https://doi.org/10.1038/nature21376).

- [79] Pierre Le Merre et al. "Reward-Based Learning Drives Rapid Sensory Signals in Medial Prefrontal Cortex and Dorsal Hippocampus Necessary for Goal-Directed Behavior". In: *Neuron* 97.1 (Jan. 2018), 83–91.e5. DOI: [10.1016/j.neuron.2017.11.031](https://doi.org/10.1016/j.neuron.2017.11.031).
- [80] Malavika Murugan et al. "Combined Social and Spatial Coding in a Descending Projection from the Prefrontal Cortex". In: *Cell* 171.7 (Dec. 2017), 1663–1677.e16. DOI: [10.1016/j.cell.2017.11.002](https://doi.org/10.1016/j.cell.2017.11.002).
- [81] Elizabeth A. Amadei et al. "Dynamic corticostriatal activity biases social bonding in monogamous female prairie voles". In: *Nature* 546.7657 (June 2017), pp. 297–301. DOI: [10.1038/nature22381](https://doi.org/10.1038/nature22381).
- [82] Caitlin M. Vander Weele et al. "Dopamine enhances signal-to-noise ratio in cortical-brainstem encoding of aversive stimuli". In: *Nature* 563.7731 (Nov. 2018), pp. 397–401. DOI: [10.1038/s41586-018-0682-1](https://doi.org/10.1038/s41586-018-0682-1).
- [83] Robert R. Rozeske et al. "Prefrontal-Periaqueductal Gray-Projecting Neurons Mediate Context Fear Discrimination". In: *Neuron* 97.4 (Feb. 2018), 898–910.e6. DOI: [10.1016/j.neuron.2017.12.044](https://doi.org/10.1016/j.neuron.2017.12.044).
- [84] Earl K. Miller et al. "An Integrative Theory of Prefrontal Cortex Function". In: *Annual Review of Neuroscience* 24.1 (2001), pp. 167–202. DOI: [10.1146/annurev.neuro.24.1.167](https://doi.org/10.1146/annurev.neuro.24.1.167).
- [85] Rajeev V. Rikhye et al. "Thalamic regulation of switching between cortical representations enables cognitive flexibility". In: *Nature Neuroscience* 21.12 (Dec. 2018), pp. 1753–1763. DOI: [10.1038/s41593-018-0269-z](https://doi.org/10.1038/s41593-018-0269-z).
- [86] Sandra Reinert et al. "Mouse prefrontal cortex represents learned rules for categorization". In: *Nature* 593.7859 (May 2021), pp. 411–417. DOI: [10.1038/s41586-021-03452-z](https://doi.org/10.1038/s41586-021-03452-z).
- [87] Liya Ma et al. "A Quantitative Analysis of Context-Dependent Remapping of Medial Frontal Cortex Neurons and Ensembles". In: *Journal of Neuroscience* 36.31 (Aug. 2016), pp. 8258–8272. DOI: [10.1523/JNEUROSCI.3176-15.2016](https://doi.org/10.1523/JNEUROSCI.3176-15.2016).
- [88] Michael J Siniscalchi et al. "Fast and slow transitions in frontal ensemble activity during flexible sensorimotor behavior". In: *Nature Neuroscience* 19.9 (Sept. 2016), pp. 1234–1242. DOI: [10.1038/nn.4342](https://doi.org/10.1038/nn.4342).
- [89] Hugo Malagon-Vina et al. "Fluid network dynamics in the prefrontal cortex during multiple strategy switching". In: *Nature Communications* 9.1 (Jan. 2018), p. 309. DOI: [10.1038/s41467-017-02764-x](https://doi.org/10.1038/s41467-017-02764-x).
- [90] Christian Bravo-Rivera et al. "Neural Structures Mediating Expression and Extinction of Platform-Mediated Avoidance". In: *Journal of Neuroscience* 34.29 (July 2014), pp. 9736–9742. DOI: [10.1523/JNEUROSCI.0191-14.2014](https://doi.org/10.1523/JNEUROSCI.0191-14.2014).
- [91] Justin M Moscarello et al. "Flexibility in the face of fear: hippocampal–prefrontal regulation of fear and avoidance". In: *Current Opinion in Behavioral Sciences. Emotion-cognition interactions* 19 (Feb. 2018), pp. 44–49. DOI: [10.1016/j.cobeha.2017.09.010](https://doi.org/10.1016/j.cobeha.2017.09.010).
- [92] Erik B. Oleson et al. "Subsecond Dopamine Release in the Nucleus Accumbens Predicts Conditioned Punishment and Its Successful Avoidance". In: *Journal of Neuroscience* 32.42 (Oct. 2012), pp. 14804–14808. DOI: [10.1523/JNEUROSCI.3087-12.2012](https://doi.org/10.1523/JNEUROSCI.3087-12.2012).

- [93] Franchesca Ramirez et al. "Active Avoidance Requires a Serial Basal Amygdala to Nucleus Accumbens Shell Circuit". In: *Journal of Neuroscience* 35.8 (Feb. 2015), pp. 3470–3477. DOI: [10.1523/JNEUROSCI.1331-14.2015](https://doi.org/10.1523/JNEUROSCI.1331-14.2015).
- [94] June-Seek Choi et al. "The role of amygdala nuclei in the expression of auditory signaled two-way active avoidance in rats". In: *Learning & Memory* 17.3 (Mar. 2010), pp. 139–147. DOI: [10.1101/lm.1676610](https://doi.org/10.1101/lm.1676610).
- [95] Jun Ma et al. "Divergent projections of the paraventricular nucleus of the thalamus mediate the selection of passive and active defensive behaviors". In: *Nature Neuroscience* 24.10 (Oct. 2021), pp. 1429–1440. DOI: [10.1038/s41593-021-00912-7](https://doi.org/10.1038/s41593-021-00912-7).
- [96] Christian Bravo-Rivera et al. "Persistent active avoidance correlates with activity in prelimbic cortex and ventral striatum". In: *Frontiers in Behavioral Neuroscience* 9 (July 2015). DOI: [10.3389/fnbeh.2015.00184](https://doi.org/10.3389/fnbeh.2015.00184).
- [97] Justin M. Moscarello et al. "Active Avoidance Learning Requires Prefrontal Suppression of Amygdala-Mediated Defensive Reactions". In: *Journal of Neuroscience* 33.9 (Feb. 2013), pp. 3815–3823. DOI: [10.1523/JNEUROSCI.2596-12.2013](https://doi.org/10.1523/JNEUROSCI.2596-12.2013).
- [98] Maria M Diehl et al. "Active avoidance requires inhibitory signaling in the rodent prelimbic prefrontal cortex". In: *eLife* 7 (May 2018), e34657. DOI: [10.7554/eLife.34657](https://doi.org/10.7554/eLife.34657).
- [99] Daniel Jercog et al. "Dynamical prefrontal population coding during defensive behaviours". In: *Nature* 595.7869 (July 2021), pp. 690–694. DOI: [10.1038/s41586-021-03726-6](https://doi.org/10.1038/s41586-021-03726-6).
- [100] Maria M Diehl et al. "Divergent projections of the prelimbic cortex bidirectionally regulate active avoidance". In: *eLife* 9 (Oct. 2020), e59281. DOI: [10.7554/eLife.59281](https://doi.org/10.7554/eLife.59281).
- [101] Tsai-Wen Chen et al. "Ultrasensitive fluorescent proteins for imaging neuronal activity". In: *Nature* 499.7458 (July 2013), pp. 295–300. DOI: [10.1038/nature12354](https://doi.org/10.1038/nature12354).
- [102] George Paxinos et al. *The Mouse Brain in Stereotaxic Coordinates*. 3rd ed. Amsterdam, Sept. 2007.
- [103] Alexander Mathis et al. "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning". In: *Nature Neuroscience* 21.9 (Sept. 2018), pp. 1281–1289. DOI: [10.1038/s41593-018-0209-y](https://doi.org/10.1038/s41593-018-0209-y).
- [104] Eftychios A. Pnevmatikakis. "Analysis pipelines for calcium imaging data". In: *Current opinion in neurobiology* 55 (2019), pp. 15–21.
- [105] Carsen Stringer et al. "Computational processing of neural recordings from calcium imaging data". In: *Current opinion in neurobiology* 55 (2019), pp. 22–31.
- [106] Eftychios A. Pnevmatikakis et al. "Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data". In: *Neuron* 89.2 (Jan. 2016), pp. 285–299. DOI: [10.1016/j.neuron.2015.11.037](https://doi.org/10.1016/j.neuron.2015.11.037).
- [107] Marius Pachitariu et al. "Suite2p: beyond 10,000 neurons with standard two-photon microscopy". In: *BioRxiv* (2017).
- [108] Pengcheng Zhou et al. "Efficient and accurate extraction of in vivo calcium signals from microendoscopic video data". In: *eLife* 7 (), e28728. DOI: [10.7554/eLife.28728](https://doi.org/10.7554/eLife.28728).

- [109] Jinghao Lu et al. "MIN1PIPE: A Miniscope 1-Photon-Based Calcium Imaging Signal Extraction Pipeline". In: *Cell Reports* 23.12 (June 2018), pp. 3673–3684. DOI: [10.1016/j.celrep.2018.05.062](https://doi.org/10.1016/j.celrep.2018.05.062).
- [110] Eran A. Mukamel et al. "Automated Analysis of Cellular Signals from Large-Scale Calcium Imaging Data". In: *Neuron* 63.6 (Sept. 2009), pp. 747–760. DOI: [10.1016/j.neuron.2009.08.009](https://doi.org/10.1016/j.neuron.2009.08.009).
- [111] P. Thevenaz et al. "A pyramid approach to subpixel registration based on intensity". In: *IEEE Transactions on Image Processing* 7.1 (Jan. 1998), pp. 27–41. DOI: [10.1109/83.650848](https://doi.org/10.1109/83.650848).
- [112] A. Hyvarinen. "Fast and robust fixed-point algorithms for independent component analysis". In: *IEEE Transactions on Neural Networks* 10.3 (May 1999), pp. 626–634. DOI: [10.1109/72.761722](https://doi.org/10.1109/72.761722).
- [113] Ethan M. Meyers et al. "Incorporation of new information into prefrontal cortical activity after learning working memory tasks". In: *Proceedings of the National Academy of Sciences* 109.12 (Mar. 2012), pp. 4651–4656. DOI: [10.1073/pnas.1201022109](https://doi.org/10.1073/pnas.1201022109).
- [114] Andrew H. Moberly et al. "Olfactory inputs modulate respiration-related rhythmic activity in the prefrontal cortex and freezing behavior". In: *Nature Communications* 9.1 (Apr. 2018), p. 1528. DOI: [10.1038/s41467-018-03988-1](https://doi.org/10.1038/s41467-018-03988-1).
- [115] Sophie Bagur et al. "Breathing-driven prefrontal oscillations regulate maintenance of conditioned-fear evoked freezing independently of initiation". In: *Nature Communications* 12.1 (May 2021), p. 2605. DOI: [10.1038/s41467-021-22798-6](https://doi.org/10.1038/s41467-021-22798-6).
- [116] Edward Zagha et al. "The Importance of Accounting for Movement When Relating Neuronal Activity to Sensory and Cognitive Processes". In: *Journal of Neuroscience* 42.8 (Feb. 2022), pp. 1375–1382. DOI: [10.1523/JNEUROSCI.1919-21.2021](https://doi.org/10.1523/JNEUROSCI.1919-21.2021).
- [117] Carsen Stringer et al. "Spontaneous behaviors drive multidimensional, brain-wide activity". In: *Science* 364.6437 (Apr. 2019), eaav7893. DOI: [10.1126/science.aav7893](https://doi.org/10.1126/science.aav7893).
- [118] Simon Musall et al. "Single-trial neural dynamics are dominated by richly varied movements". In: *Nature Neuroscience* 22.10 (Oct. 2019), pp. 1677–1686. DOI: [10.1038/s41593-019-0502-4](https://doi.org/10.1038/s41593-019-0502-4).
- [119] Nicholas A. Steinmetz et al. "Distributed coding of choice, action and engagement across the mouse brain". In: *Nature* 576.7786 (Dec. 2019), pp. 266–273. DOI: [10.1038/s41586-019-1787-x](https://doi.org/10.1038/s41586-019-1787-x).
- [120] Bridget L. Kajs et al. *Divergent encoding of active avoidance behavior in corticostriatal and corticolimbic projections*. Tech. rep. bioRxiv, Sept. 2021, p. 2021.09.15.460552. DOI: [10.1101/2021.09.15.460552](https://doi.org/10.1101/2021.09.15.460552).
- [121] Lisa A. Gunaydin et al. "Cortico-Basal Ganglia Circuit Function in Psychiatric Disease". In: *Annual Review of Physiology* 78 (2016), pp. 327–350. DOI: [10.1146/annurev-physiol-021115-105355](https://doi.org/10.1146/annurev-physiol-021115-105355).
- [122] Michael S Fanselow. "Emotion, motivation and function". In: *Current Opinion in Behavioral Sciences*. Emotion-cognition interactions 19 (Feb. 2018), pp. 105–109. DOI: [10.1016/j.cobeha.2017.12.013](https://doi.org/10.1016/j.cobeha.2017.12.013).

Chapter 7

Appendix

Bleaching model for preprocessing of calcium imaging movies

As described in section 2.3.1, we observed issues with bleaching in our calcium imaging movies, mainly reflected by a general decrease in signal intensity (Fig. 7.1, top row). However, not all pixels were equally affected and particularly pixels related to blood vessel patterns showed different changes than other pixels (Fig. 7.1, bottom right). This non-uniformity precluded dealing with bleaching using simple filtering techniques. Instead, we constructed a rank-2 bleaching model using PCA as illustrated in Fig. 7.2. The two first principle components in pixel space corresponded to the two differently behaving pixel patterns described above: PC1 reflected changes with a low spatial frequency, whereas PC2 reflected blood vessel patterns. These two patterns had opposing temporal dynamics where PC1 intensity decreased while PC2 intensity increased (Fig. 7.2, bottom right). Importantly, bleaching models were stable over sessions of different days and thus did not introduce variability between recording sessions.

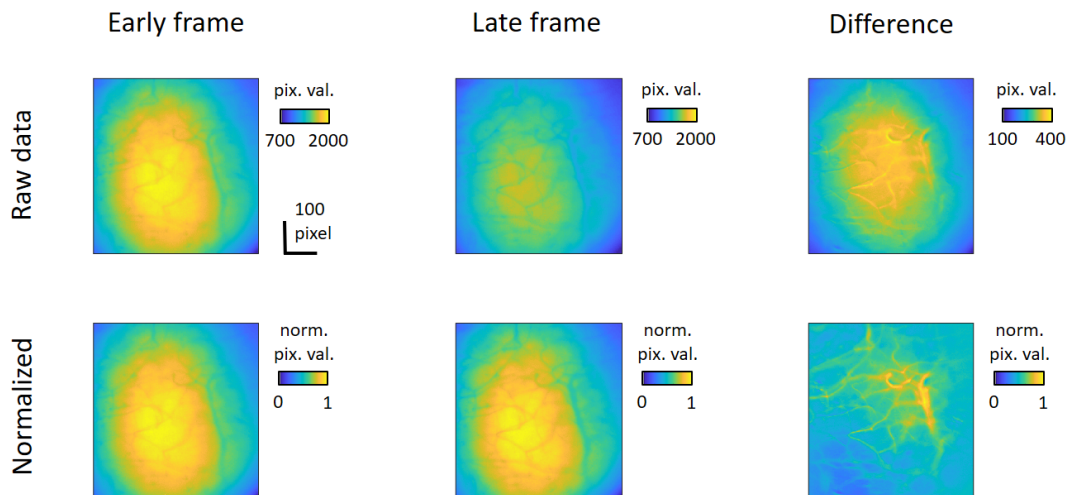


FIGURE 7.1: Illustration of non-uniform bleaching. Top row: (left) Raw example frame from the beginning of an example session. (middle) Example frame from late in the same session. (right) Difference between early and late frame. Bottom row: (left, middle) Normalized frames. (right) Difference of normalized frames highlighting non-uniform changes related to blood vessel patterns.

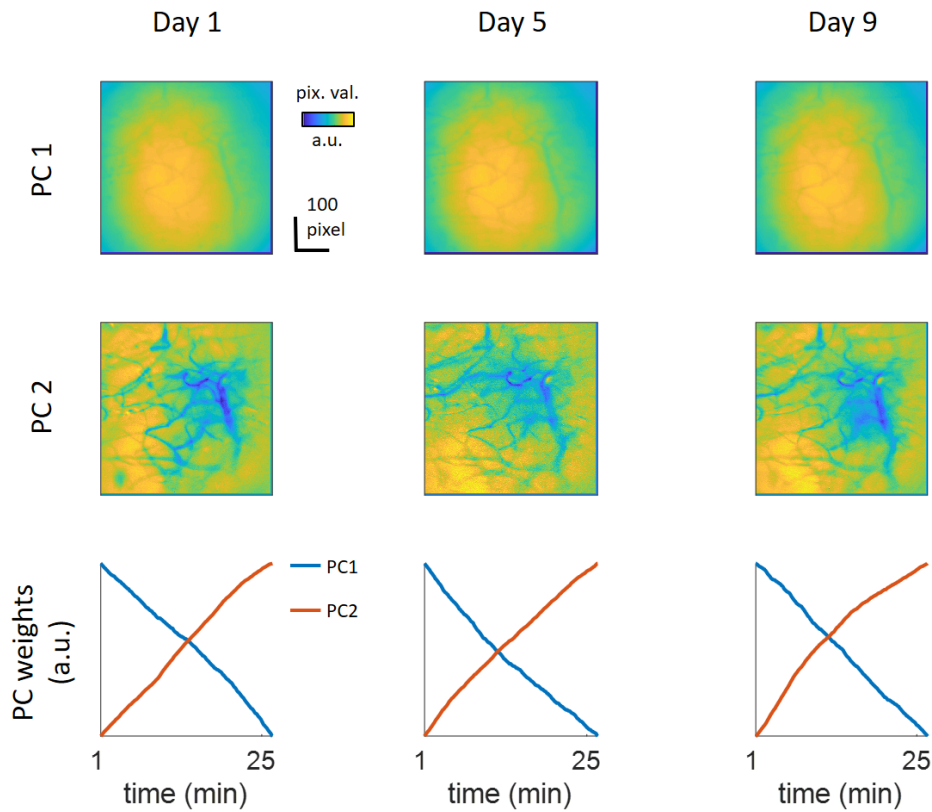


FIGURE 7.2: Illustration of bleaching models on different days. Top row: PC1 in pixel space corresponding to a pixel pattern with low spatial frequency which decreases intensity over recording sessions. Middle row: PC in pixel space corresponding to a blood vessel pattern that increases in intensity relative to PC1. Bottom row: Temporal weights of the two PCs highlighting the opposing temporal evolution.

Example cells from signal extraction annotation

Here we display further example cells from the annotation process that we use to validate the automatically extracted cells as described in section 2.3.5. The first two cells represent two further positive examples, while the remaining cells display characteristics that led to their exclusion as described in the captions.

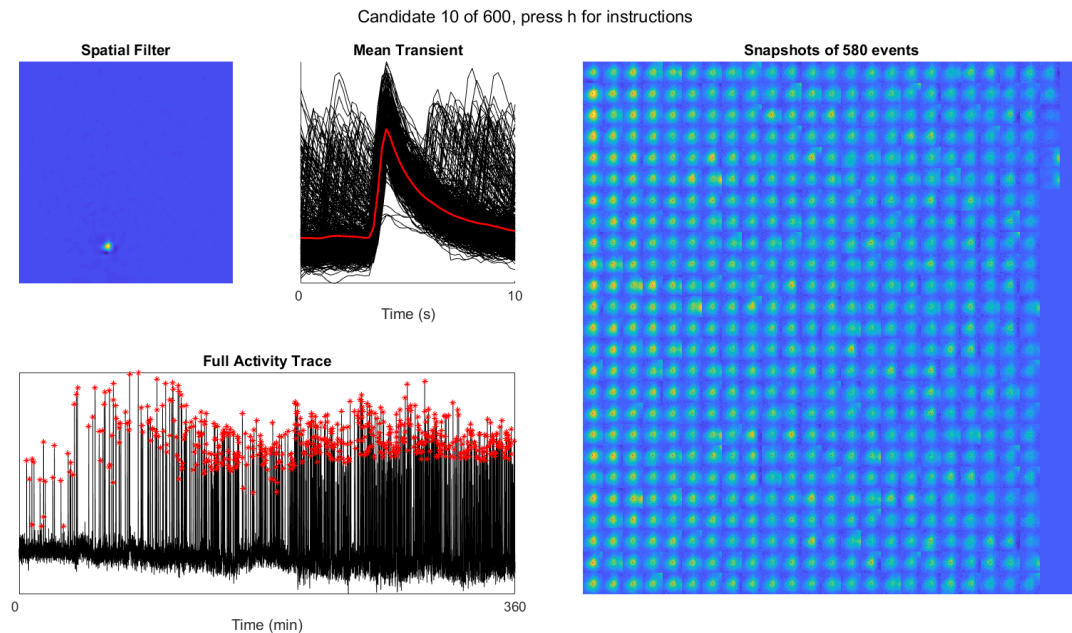


FIGURE 7.3: Example cell that was accepted in the annotation process.

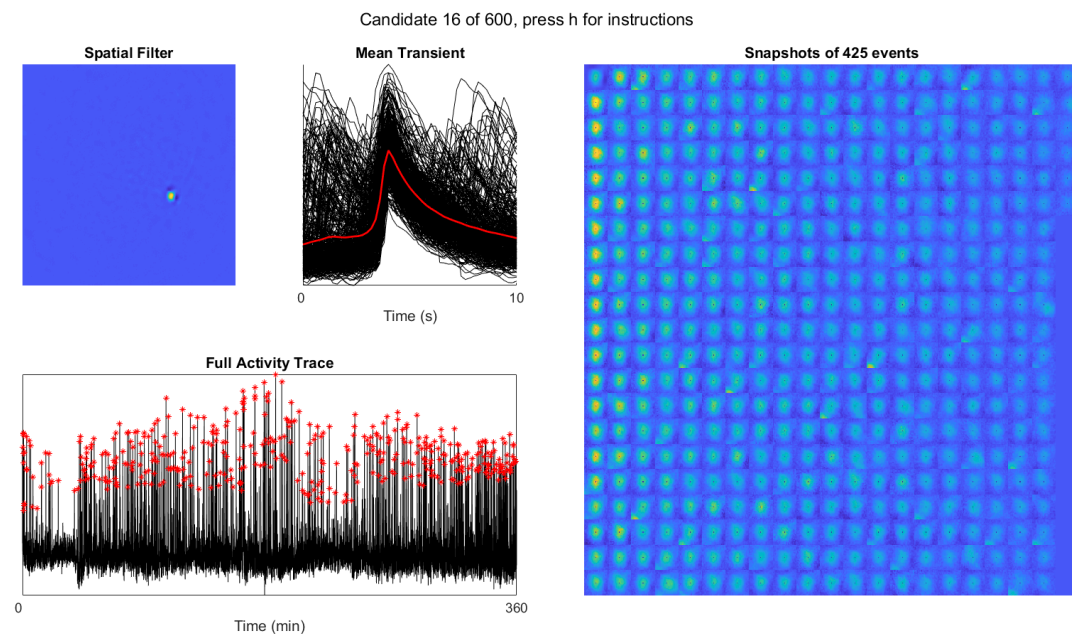


FIGURE 7.4: Example cell that was accepted in the annotation process.

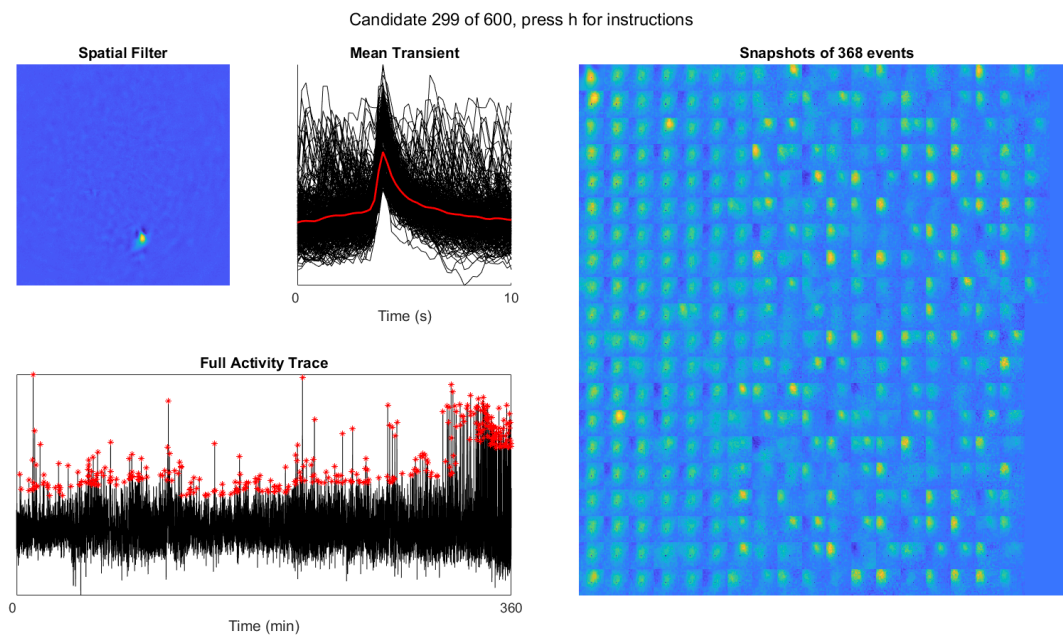


FIGURE 7.5: Example cell that was rejected in the annotation process because of cross talk between two overlapping cells. This cross talk can be seen by the difference between event snapshots.

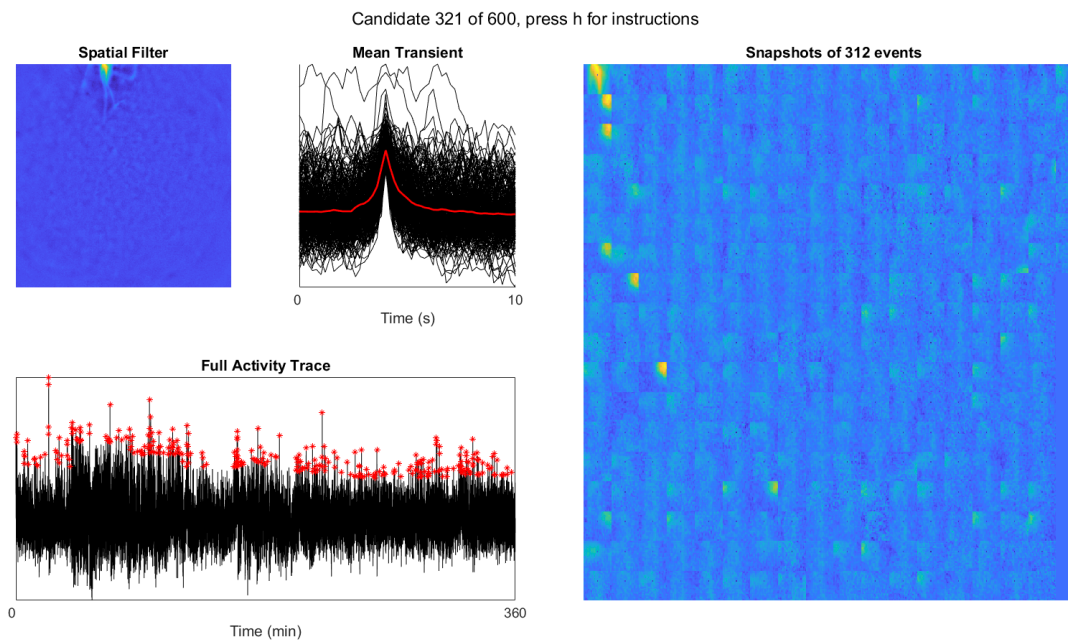


FIGURE 7.6: Example cell that was rejected in the annotation process because of untypical shape, untypical temporal dynamics and differences between event snapshots.

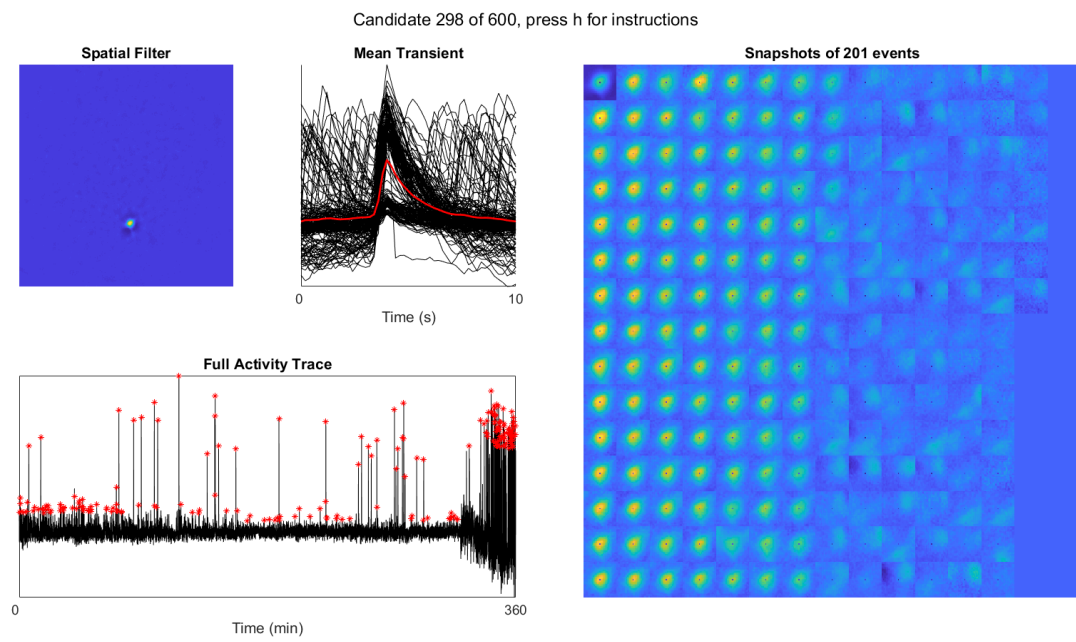


FIGURE 7.7: Example cell that was rejected in the annotation process because of lacking stability over sessions. The noisy period towards the end corresponds to the final recording session.

Further quantification of task specificity of 2DAA tone responses

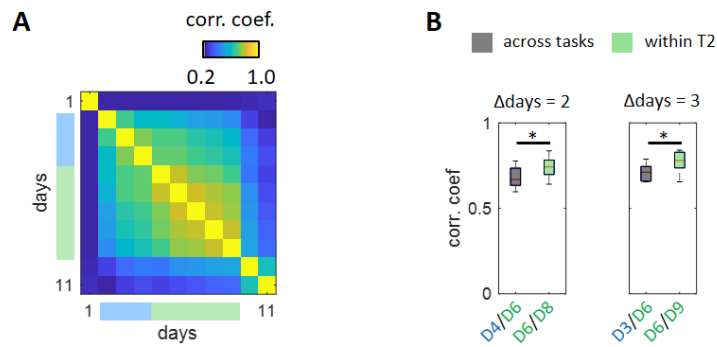


FIGURE 7.8: Quantification of task-specificity of tone responses using z-score correlations. **(A)** Pairwise Pearson correlation coefficients of tone z-scores per day. Mean over 12 subjects. **(B)** Quantification of task-specificity of tone responses comparing within-task and across-task similarity at two different time shifts. Wilcoxon signed rank test ($p < 0.05$), $n = 12$ subjects.

Contributions

Signal extraction code

- Written by Benjamin Ehret, adapted from a code base from the Schnitzer Lab in Stanford

Fear conditioning

- All experiments were performed by Benjamin Grewe in the Schnitzer Lab in Stanford
- All data analysis except for DLC tracking was performed by Benjamin Ehret
- DLC tracking was by Roman Boehringer

Active avoidance

- The active avoidance setup was built by Christian Henning and Benjamin Ehret
- Surgeries (viral injections and microendoscope implantations) were performed by Roman Boehringer (70/120) and Benjamin Grewe (10/120) and Benjamin Ehret (40/120)
- All microscope mountings were performed by Benjamin Ehret
- Imaging sessions during behavior were recorded by Roman Boehringer (8/287) and Benjamin Ehret (279/287)
- All perfusions and histological analyses were performed by Roman Boehringer
- All data analysis except for DLC tracking was performed by Benjamin Ehret
- DLC tracking was by Roman Boehringer

Publications

- **Ehret B.***, Henning C.*, Cervera M. R.*, Meulemans A., von Oswald J. and Grewe B. F. *Continual Learning in Recurrent Neural Networks*. International Conference on Learning Representations, 2021
- Henning C.*, Cervera M.R.*, D'Angelo F., von Oswald J., Traber R., **Ehret B.**, Kobayashi S., Grewe B.F. and Sacramento J. *Posterior Meta-Replay for Continual Learning*. Conference on Neural Information Processing Systems, 2021
- Duering D.N., Dittrich F., Rocha M. D., Tachibana R. O., Mori C., Okanoya K., Boehringer R., **Ehret B.**, Grewe B. F., Gerber S., Ma S., Rauch M., Paterna J., Kasper R., Gahr M. and Hahnloser R. H. R. *Fast retrograde access to projection neuron circuits underlying vocal learning in songbirds*. Cell Reports, 2020

Curriculum Vitae

Personal data

Name	Benjamin Ehret
Date of Birth	March 06, 1990
Place of Birth	München, Germany
Citizen of	Germany

Education

2016 – Present	Institute of Neuroinformatics University of Zürich and ETH Zürich Zürich, Switzerland <i>Final degree: PhD</i>
2014 – 2016	University of Zürich and ETH Zürich Zürich, Switzerland <i>Final degree: MSc in Neural Systems and Computation</i>
2010 – 2014	LMU München and TU München München, Germany <i>Final degree: BSc in Bioinformatics</i>
2012	Université Laval Québec City, Canada <i>Semester Abroad</i>
2000 – 2009	Feodor-Lynen-Gymnasium Planegg, Germany <i>Final degree: Abitur (university entrance diploma)</i>