

**Multidimensional Coding by Neurons of the Basolateral Amygdala
and Prefrontal Cortex**

By: Pinelopi Kyriazi

A dissertation submitted to the Graduate School – Newark

Rutgers, The State University of New Jersey

in partial fulfillment of requirements for the degree of

Doctor of Philosophy

Graduate Program in Behavioral and Neural Sciences

Written under the direction of

Dr. Denis Pare

and approved by

Tibor Koos, Ph.D.

Bart Krekelberg, Ph.D.

Vincent McGinty, Ph.D.

Denis Pare, Ph.D.

Daniel Salzman, M.D., Ph.D.

October, 2020

© 2020

Pinelopi Kyriazi

ALL RIGHTS RESERVED

Abstract of the Thesis

Multidimensional Coding by Neurons of the Basolateral Amygdala
and Prefrontal Cortex

By: Pinelopi Kyriazi

Dissertation Director: Dr. Denis Pare

The medial prefrontal cortex (mPFC) and basolateral complex of the amygdala (BLA), consisting of the lateral (LA), and basolateral (BL) and basomedial nuclei, are two nodes involved in the regulation of emotional memories. Specifically, BLA is involved in the acquisition and expression of conditioned responses (CRs) to stimuli (CSs) that predict aversive or appetitive outcomes. These CRs are thought to result from the activation of specific subsets of valence-coding BLA neurons. Under this model, the responses of BLA cells to CSs and the activity that drives CRs are closely related. In the third chapter of this thesis, I tested this hypothesis using a novel paradigm that dissociates BLA activity related to CS responses and CRs. At odds with this model, the CS responses and CR-related activity of individual BLA cells were separable. Moreover, while the incidence of valence-coding cells did not exceed chance, at the population level there was similarity between valence coding for CSs and CRs. In fact, both LA and BL neurons concurrently encoded multiple task features and behaviors.

The mPFC forms dense reciprocal connections with the BLA and has been shown to serve similar functions as the BLA in fear and reward conditioning. Additionally, these

two brain regions have common cortical, striatal, and brainstem connections involved in emotional expression, neuromodulation, and autonomic regulation. Hence, to understand the overlapping and separable roles of the mPFC and BLA in appetitive and aversive conditioning, in the fourth chapter of this thesis, I recorded cells in both regions and compared their coding properties. Several similarities were apparent, including the proportions of stimulus and behavior responsive cells. However, mPFC neurons had more selective coding especially in relation to CRs evoked by different CSs. Importantly, neurons in the two structures encoded the same task features but to different degrees, leading to gradients of task representations. While PL and BL neurons represented many task variables through small fluctuations in their firing rates, LA cells coded fewer task features with stronger changes in activity. Population analyses revealed that while valence could be decoded from the activity of BLA and PL neurons, PL neurons were superior at distinguishing trial identity information. Thus, PL neurons have greater representational capacity than BLA neurons. The significance of these differences for the regulation of emotional memories is discussed.

Acknowledgements

This journey could not have been possible without the support of my family, friends, and mentors. First, I want to thank my incredible mentors, Drs. Denis Paré and Drew Headley, who have been vital to my growth and development as a scientist. Through our interactions, I have advanced my scientific understanding, I have been challenged to think more critically, and most importantly grown into a professional scientist. Additionally, I want to thank my thesis committee and program director, Drs. Vincent McGinty, Bart Krekelberg, Tibor Koos, Daniel Salzman, and Joan Morrell for your input and guidance throughout this process.

Second, I want to thank my family, fiancé, and friends for being there when I needed you the most. Especially my mom, dad, and sister for supporting me throughout my life and providing all the resources I needed to succeed mentally, emotionally, and professionally. Mom and dad, thank you for making the biggest decision of your lives to move to the United States so I can have moments like this and a bright future. Also, thank you to my family in Greece for always being there for me, visiting us when possible, and spending time with us every summer. Most importantly, a big thank you to my fiancé, Thimios Lironis, who moved states to help me finish this journey, listened to countless hours of my science presentations, and was always happy to engage in scientific debates with me. My friends and labmates Vicky, Samar, Sally, Ian, John, Hussein, Luisa, Alon, Zeynep, and countless others who made graduate school fun and pushed me to try new things.

Finally, thank you to the CMBN administrative and animal facility staff. It was a pleasure working with you to plan department events and give back to this community. A

big thank you to Felipe in the animal facility, who always took great care of my animals and genuinely cared for my wellbeing. My time at Rutgers has been amazing and this experience has shaped me into the person I am today. I am grateful for all the people I met, the things I learned, and the journey I experienced.

Table of Contents

Abstract of the Thesis	ii
Acknowledgements.....	iv
Table of Contents.....	vi
List of Illustrations	ix
Chapter I: Introduction.....	1
1.1 Defensive behaviors are shared across mammalian species.....	1
1.2 Role of the basolateral amygdala in conditioned emotional behaviors.....	4
1.3 A novel paradigm for studying defensive and appetitive behaviors	6
1.4 Role of the medial prefrontal cortex in conditioned emotional behaviors	7
1.5 Comparing the role of BLA and mPFC in conditioned emotional behaviors	10
Chapter II:	14
General methods	14
2.1 Subjects.....	14
2.2 Risk Reward Interaction Task	14
2.3 Surgical procedures	16
2.4 Unit recordings and clustering.....	19
2.5 Histology	21
2.6 Behavioral Analyses.....	22
2.7 Classification of valence neurons	22
2.8 Generalized Linear Model (GLM)	23
2.9 Multidimensional scaling (MDS)	26
2.10 Normalized peak firing rate modulations.	27

2.11 Similarity Matrices	28
2.12 Number of Features Encoded.	28
2.13 Gini and dimensionality indices	29
2.14 D-prime analyses	30
2.15 Principal Component Analysis (PCA).....	31
2.16 Support Vector Machine Decoder	32
2.17 Statistical Analyses.....	33
Chapter III:.....	35
Multidimensional coding by basolateral amygdala neurons.....	35
3.1 Rationale.....	35
3.2 Rats quickly learn the Risk Reward Interaction task.....	35
3.3 Activity of principal BLA cells during CSs	37
3.4 Activity of principal BLA cells in relation to CRs.....	40
3.5 Generalized linear model untangles cell activity.....	43
3.6 Task-related activity in the BLA and striatum	46
3.7 Heterogeneous Coding in LA and BL	52
3.8 Conclusions	58
Chapter IV:.....	60
Multidimensional representations in the amygdalo-prefrontal network.....	60
4.1 Rationale.....	60
4.2 CS responsiveness and behavioral correlates of PL neurons	61
4.3 Differences between the selectivity of PL and BLA neurons	67
4.4 Disentangling the correlates of unit activity with a generalized linear model (GLM).....	69
4.5 Population analyses reveal multidimensional representations	76

4.6 Conclusions	83
Chapter V:	85
General Discussion	85
5.1 Independent encoding of CSs and CRs by BLA neurons.....	85
5.2 At the level of individual cells, valence-coding is not prevalent in the BLA	86
5.3 Multidimensional coding by BLA neurons	88
5.4 Comparing BLA and mPFC activity	89
5.5 Gradations in task representations.....	90
5.6 Gradations in the manner of population representations	93
5.7 Codes and computations.....	95
5.8 Gaps and future directions.....	95
List of references:	98

List of Illustrations

Figure 1. The rodent amygdala.	3
Figure 2. The rodent medial prefrontal cortex.	8
Figure 3. Risk-reward interaction task.	16
Figure 4. Methods used to record, classify, cluster, and locate BLA neurons.	18
Figure 5. GLM basis and kernel functions.	26
Figure 6. Incidence, duration, and latency of different types of CRs during training.	37
Figure 7. Activity of BLA neurons during the CS-Rs and CS-Ss.	38
Figure 8. Behavioral correlates of unit activity.	41
Figure 9. Coding of task variables by example BLA cells, as estimated by the GLM.	44
Figure 10. Coefficient of determination for principal cells and interneurons.	45
Figure 11. Multidimensional coding by BLA neurons, as determined by the GLM.	48
Figure 12. Relation between GLM-estimated coding of task variables and firing rate change observed during active conditioned behaviors.	51
Figure 13. Correlation between firing rate modulations associated with behaviors and CSs.	53
Figure 14. Coding for different task dimensions at the population level.	55

Figure 15. Putative interneurons in both the amygdala and striatum exhibited systematic mapping of coding.	58
Figure 16. RRI task, histological verification of PL neurons, and their classification	62
Figure 17. CS-evoked activity in LA, BL, and PL neurons.....	63
Figure 18. CS and behavior response patterns of LA, BL, and PL neurons.	64
Figure 19. CS- and CR-related activity of PL neurons.	66
Figure 20. Higher incidence of CS- and CR-selective cells in PL.....	68
Figure 21. Coding of task variables by an example PL cell as estimated by the GLM....	70
Figure 22. Example PL neuron with speed correlate fit with GLM and model accuracy.	71
Figure 23. Comparison between coding of RRI task variables by striatal, LA, BL, and PL neurons, as determined using the GLM.	73
Figure 24. Contrasting representation of task variables in the striatum, amygdala, and prefrontal cortex.	74
Figure 25. D-prime values.	75
Figure 26. Principal component analysis trajectories by trial type.	77
Figure 27. Neural representation of position based on the GLM.	78
Figure 28. Trial type and valence decoders.	80

Figure 29. Cross-temporal generalization of the SVM decoder. 83

Figure 30. Representation gradients across the amygdalo-striatal-prefrontal network. ... 91

Chapter I:

Introduction

1.1 Defensive behaviors are shared across mammalian species

Animals respond to threats using a diverse repertoire of defensive behaviors that promote survival. These behaviors have been retained by natural selection and are expressed similarly in various mammalian species, including humans. Indeed, while most studies of defensive behaviors have been performed in rodents, their results parallel those obtained in monkeys and humans (Blanchard, Hynd, Minke, Minemoto, & Blanchard, 2001; Kalin & Shelton, 1989). Defensive behaviors commonly studied in rodents include freezing (to avoid detection), fighting, fleeing, making vocalizations (that warn conspecifics a threat is present), and quickly orienting toward unexpected stimuli (Headley, Kanta, Kyriazi, & Pare, 2019). Environmental factors such as predator distance and the availability of escape routes play a major role in determining the type of defensive behavior evoked by threats (Roelofs, 2017). Distal threats in environments that lack escape routes evoke freezing. When escape options are available, proximal and imminent threats lead to fleeing or avoidance. In contrast, when there are no opportunities for escape, they elicit fighting (Blanchard, Griebel, Pobbe, & Blanchard, 2011; Fanselow, 1994; Perusini & Fanselow, 2015).

The neural circuits mediating defensive behaviors are also similar across the mammalian kingdom. Early studies in monkeys identified the medial temporal lobe as a critical brain region for fear responses. When the medial temporal lobe was lesioned, monkeys showed reduced fear responses to aversive stimuli (Kluver & Bucy, 1997). Most importantly, it was later determined that this impairment was due to lesioning of the amygdala (Weiskrantz, 1956). Additionally, human neuroimaging studies of fear conditioning have implicated similar amygdala pathways as seen in rodents and monkeys (Mobbs et al., 2010; Phelps & LeDoux, 2005). Given this evidence, many studies have focused on the role of the amygdala in fear learning and expression.

The amygdala is a medial temporal lobe structure consisting of several nuclei and cortex-like structures (Pape & Pare, 2010), which can be distinguished based on their connections as well as cyto- and chemo-architectonic features (Pitkanen, Savander, & LeDoux, 1997). Three main amygdala nuclei are involved in regulating conditioned fear and emotional memories (**Fig. 1**): the lateral nucleus (LA), the basolateral nucleus (BL), and the central nucleus (CE). The LA, BL, and basomedial nuclei together form the basolateral complex of the amygdala (BLA). Morphologically, the BLA has a similar neuronal composition as the cerebral cortex, with the exception that the latter has a laminar organization. In contrast, BLA cells appear to be oriented randomly throughout the structure, with approximately 80% being glutamatergic pyramidal neurons and 20% GABAergic (Muller, Mascagni, & McDonald, 2006; Pape & Pare, 2010).

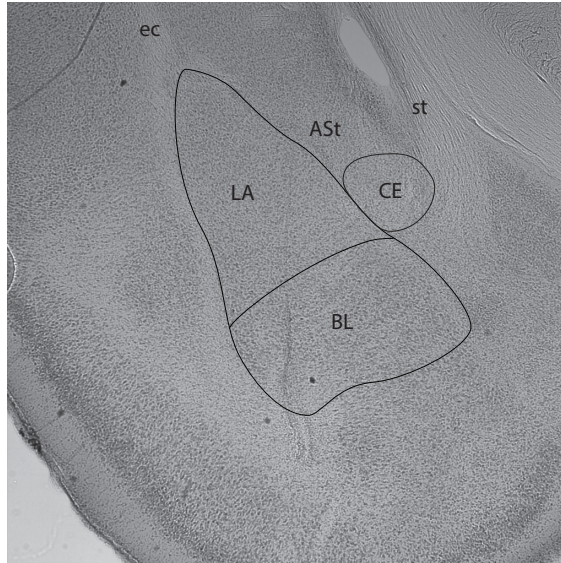


Figure 1. The rodent amygdala.

Thionin-stained coronal section of the rat amygdala. The lateral nucleus (LA), basolateral nucleus (BL), and central nucleus (CE) are outlined. ec: external capsule; ASt: amygdalostriatal transition area; st: stria terminalis.

In general terms, most sensory information enters the amygdala through LA. Tracing studies have indicated that auditory and visual sensory inputs from the thalamus and sensory association cortices mainly target the dorsolateral portion of LA (Doron & LeDoux, 1999; LeDoux, Farb, & Ruggiero, 1990; Romanski, Clugnet, Bordi, & LeDoux, 1993), albeit BL and CE also receive some sensory inputs. In turn, the dorsolateral sector of the LA projects more ventrally in the nucleus, which then projects to BL (Pitkanen et al., 1997). Unlike LA, BL has dense rostrocaudal connections within itself, allowing for intradivisional communication (Pitkanen et al., 1997). Additionally, BL is reciprocally connected with cortical structures including the medial prefrontal cortex (mPFC), the insula, perirhinal and entorhinal cortices, and various hippocampal fields (Pape & Pare, 2010). While LA is also connected with cortical structures, particularly the perirhinal cortex, LA's cortical connections are typically less dense than

BL's, especially in relation to mPFC (Hintiryan et al., 2019; McDonald, 1991; McDonald, Mascagni, & Guo, 1996).

From the BLA, information flows to CE via unidirectional projections (Pitkanen et al., 1997). CE is considered the major output of the amygdala to fear effector structures, such as the brainstem and hypothalamus. CE is a striatal-like structure because it contains one main cell type: GABAergic neurons (McDonald, 1992; Pare & Smith, 1993). It is typically divided into lateral (CeL) and medial (CeM) sectors. CeM mediates freezing behaviors via its projections to the ventrolateral periaqueductal gray (Ciocchi et al., 2010; Haubensak et al., 2010). CeL is thought to gate this output because it receives excitatory input from the BLA and sends GABAergic projections to CeM. However, BL also has direct projections to medial CE, allowing for direct control of CeM output projections (Pape & Pare, 2010; Savander, Go, LeDoux, & Pitkanen, 1995).

In the following sections, I review studies that have advanced our understanding of the role of the amygdala in conditioned emotional behaviors (both aversive and appetitive). Next, I introduce a novel behavioral paradigm that overcomes the limitations of commonly used conditioning paradigms. Then, I explore the role of mPFC in mediating defensive and appetitive behaviors. Finally, I compare the role of the BLA and mPFC in the regulation of emotional memories.

1.2 Role of the basolateral amygdala in conditioned emotional behaviors.

In animals, fear has predominantly been studied using the Pavlovian fear conditioning paradigm. In this task, a conditioned stimulus (CS) such as a tone predicts the delivery of an aversive unconditioned stimulus (US), typically a footshock. As

animals learn the association between the CS and the US, they develop a freezing conditioned response (CR) to the CS. This experimental paradigm has been very useful for studying brain regions involved in the acquisition and expression of conditioned fear. However, it suffers from an important pitfall. Animals can only display one defensive behavior, freezing. Hence, our understanding of the neural circuits of fear have been constrained by studying one defensive behavior while leaving out the rest. This limitation, has led to the notion that fearful stimuli automatically activate defensive CRs. New paradigms are needed to expand our understanding of how mammals select different defensive behaviors, giving a more complete picture of how the neural circuits of fear function.

The BLA is necessary for the acquisition and expression of CRs to stimuli that predict aversive or rewarding outcomes (Ambroggi, Ishikawa, Fields, & Nicola, 2008; Fanselow & Poulos, 2005; LeDoux, 2000). In the LA, the main input station of the amygdala for sensory inputs (LeDoux, 2000; Sah, Faber, Lopez De Armentia, & Power, 2003), conditioning enhances the efficacy of synapses that convey CS information (McKernan & Shinnick-Gallagher, 1997; Rumpel, LeDoux, Zador, & Malinow, 2005; Tye, Stuber, de Ridder, Bonci, & Janak, 2008), resulting in an increased CS responsiveness of target neurons (Genuit-Gabai, Klavir, & Paz, 2013; Ghosh & Chattarji, 2015; Maren & Quirk, 2004; Quirk, Repa, & LeDoux, 1995). Within the LA and BL, largely separate populations of cells acquire responses to appetitive or aversive stimuli (Belova, Paton, & Salzman, 2008; Burgos-Robles et al., 2017; Lee, Amir, Haufler, & Pare, 2017; Sangha, Chadick, & Janak, 2013; however see Shabel & Janak, 2009) and they project to different downstream structures, which in turn mediate distinct behavioral

responses (Beyeler et al., 2016; Namburi et al., 2015). Together these findings suggest that the changes in CS responsiveness displayed by BLA cells following appetitive or aversive conditioning constitute potentiated sensory responses that can automatically trigger approach or defensive CRs through distinct neuronal outputs.

However, several observations are inconsistent with this rigid serial mechanism. For instance, BLA cells can be recruited into memory traces based on increased excitability irrespective of their underlying function (Han et al., 2007; Yiu et al., 2014). Furthermore, a recent Ca^{2+} imaging study reported that, at the population level, fear conditioning-induced changes in CS representation do not involve a potentiation of BLA responses to the CS, but a qualitative shift toward the representation of the US (Grewe et al., 2017). Yet another study revealed that the activity of BL neurons is not related to the sensory properties of the CSs but is closely tied to the reward-seeking behavior the CSs elicit (Lee, Amir, Headley, Haufler, & Pare, 2016). Last, animals display different defensive behaviors (Pellman & Kim, 2016) depending on threat proximity (Perusini & Fanselow, 2015) and a strict serial mechanism cannot account for such flexibility.

1.3 A novel paradigm for studying defensive and appetitive behaviors

As the previous studies highlight, current conditioning paradigms, such as Pavlovian fear conditioning, typically limit animal behavior to a single CR (e.g. freezing). This can give the impression that CSs automatically trigger CRs when in fact it is the only option the experimental conditions allow. Since conditioning increases the likelihood that CSs will elicit a particular CR, it is difficult to disentangle whether

training-induced alterations in BLA activity are related to the valence or sensory properties of the CSs, the behavior they elicit, or a mixture thereof.

To overcome these limitations, we developed a novel paradigm, the Risk-Reward Interaction (RRI) task, where rats can respond to a light CS in different ways depending on where the CS is presented. Rats learn that in some positions, the CS signals reward availability (CS-R) and in others, an impending footshock (CS-S). The footshock can be avoided passively or actively, depending on the rats' position in relation to the CS. Behavioral freezing is also observed, allowing us to compare neuronal activity related to different defensive behaviors. In appetitive trials, animals must approach a port to retrieve a water reward. Hence, this paradigm builds on previous tasks and allows animals to display a more diverse repertoire of behaviors. In Chapter III, I use this new paradigm to test whether the responses of BLA cells to CSs and the activity that drives CRs are closely related.

1.4 Role of the medial prefrontal cortex in conditioned emotional behaviors

The RRI task requires animals to flexibly choose between different defensive or appetitive behavioral outputs based on the location of the CS on each trial. Given this need for flexibility and decision making, the mPFC is a likely candidate area for mediating such decisions. Indeed, the mPFC has been implicated in goal-directed actions, memory and decision making, maintaining task contingencies in working memory, fear or reward-guided learning, and many other complex cognitive functions (Euston, Gruber, & McNaughton, 2012; Miller, 2000; Miller & Cohen, 2001). Furthermore, this brain region is interconnected with sensory and motor cortical areas as well as subcortical

structures such as the amygdala and hippocampus (Heidbreder & Groenewegen, 2003; McDonald, 1998; Vertes, 2004). It is widely believed that this extremely diverse connectivity allows the mPFC to process the wide range of information needed for the regulation of complex behaviors.

Prefrontal areas are mainly defined by their thalamocortical, subcortical, and cortico-cortical projections. The mPFC is subdivided into two main regions, the prelimbic (PL) and infralimbic (IL) areas (**Fig. 2**). In this thesis, I will focus on PL, which has prominent reciprocal connections with the BLA (Hoover & Vertes, 2007; McDonald, 1998; Vertes, 2004). Specifically, it was shown that the same amygdala cells that receive PL inputs project back to PL, and that PL cells receiving amygdala inputs also project to the amygdala. This connectivity allows for closed-loop routing between the two regions to exchange information (Little & Carter, 2013; McGarry & Carter, 2017).

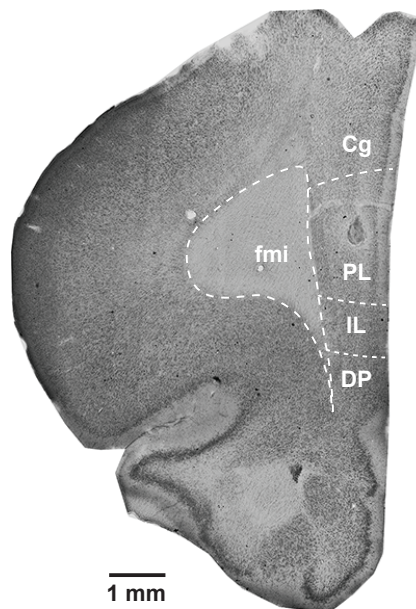


Figure 2. The rodent medial prefrontal cortex.

Thionin-stained coronal section of the rat medial prefrontal cortex. The prelimbic (PL) and infralimbic (IL) subregions are outlined. Cg: cingulate cortex; DP: dorsal peduncular cortex; fmi: forceps minor of the corpus callosum.

PL also makes cortico-cortical connections within itself and with other medial prefrontal areas including the anterior cingulate cortex, the medial and ventral orbital cortex, and the infralimbic area (IL; Hoover & Vertes, 2007). Additionally, some of these projections are contralateral, especially with the anterior PL and medial orbital areas. Other prominent cortical connections include links with the agranular insular cortex, the claustrum, as well as the perirhinal and entorhinal cortices (Hoover & Vertes, 2007; Ongur & Price, 2000; Swanson, 1981; Vertes, 2004).

Similar to other cortical areas, PL receives thalamic inputs, mainly arising from the mediodorsal nucleus (MD), the paraventricular nucleus (PVT), and nucleus reuniens. The connection with MD is reciprocal. Moreover, MD receives input from the amygdala, forming an indirect connection between the amygdala and PL (Ray & Price, 1992). PVT also projects to the amygdala, allowing it to influence both PL and BLA (Vertes & Hoover, 2008). Finally, PL receives a unidirectional projection from the hippocampus, mainly its ventral portion, and a light input from the lateral hypothalamus (Hoover & Vertes, 2007).

On the other hand, PL sends a prominent projection to nucleus accumbens (Sesack, Deutch, Roth, & Bunney, 1989), which is thought to support the role of PL in mediating appetitive behaviors (Burgos-Robles, Bravo-Rivera, & Quirk, 2013; Otis et al., 2017). Additionally, PL projects to brainstem nuclei, including the periaqueductal gray region (PAG; Floyd, Price, Ferry, Keay, & Bandler, 2000), which has been implicated in threat reactions such as freezing and escape behaviors (Amorapanth, Nader, & LeDoux, 1999; Bittencourt, Carobrez, Zamprogno, Tufik, & Schenberg, 2004; Mobbs et al., 2010).

Functionally, PL has been shown to influence both aversive and rewarding CRs. Thus, depending on the task, manipulations that enhance PL excitability will enhance drug-seeking or behavioral freezing (Giustino & Maren, 2015; Peters, Kalivas, & Quirk, 2009; Sotres-Bayon & Quirk, 2010). In a fear conditioning paradigm, the unit responses of PL neurons to conditioned tones were sustained and correlated with the expression of freezing (Burgos-Robles, Vidal-Gonzalez, & Quirk, 2009). Moreover, pharmacological inactivation of PL reduced the expression of conditioned freezing (Corcoran & Quirk, 2007; Laurent & Westbrook, 2009; D. Sierra-Mercado, Padilla-Coreano, & Quirk, 2011).

In agreement with these studies, PL is also involved in the acquisition and expression of avoidance behaviors. PL inactivation impairs the expression of avoidance (Beck et al., 2014; Bravo-Rivera, Roman-Ortiz, Brignoni-Perez, Sotres-Bayon, & Quirk, 2014), and PL activity has been associated with persistent avoidance in rats that failed to extinguish the avoidance response (Bravo-Rivera, Roman-Ortiz, Montesinos-Cartagena, & Quirk, 2015). Surprisingly however, PL recordings indicated that more neurons are inhibited during avoidance compared to classical fear conditioning, and if this inhibition is countered with optogenetic activation to maintain cells at their baseline firing, then avoidance expression is impaired (Diehl et al., 2018). These results suggest that PL interacts with the BLA during both active and passive response strategies in appetitive and aversive situations.

1.5 Comparing the role of BLA and mPFC in conditioned emotional behaviors

Of the structures regulating memory and emotions, the amygdala and mPFC stand out for their remarkably similar connectivity (reviewed in Ongur & Price, 2000; Pitkanen

et al., 1997). Not only do these structures form dense reciprocal connections with each other, but they both have access to high order sensory information, share close ties with the insula, project to overlapping striatal territories, and target a common array of brainstem nuclei involved in neuromodulation, emotional expression, and autonomic regulation (Allen, Saper, Hurley, & Cechetto, 1991; Gabbott, Warner, Jays, Salway, & Busby, 2005; Hoover & Vertes, 2007; Krettek & Price, 1977, 1978; McDonald, 1998; Ongur & Price, 2000).

Consistent with their similar connectivity, the functions of the amygdala and mPFC are tightly intertwined. For instance, they are both required for the expression and extinction of CRs to cues that predict aversive outcomes such as behavioral freezing and active avoidance (Bravo-Rivera et al., 2014; Fanselow & Poulos, 2005; LeDoux, 2000; Moscarello & LeDoux, 2013; D. Sierra-Mercado, Jr., Corcoran, Lebron-Milad, & Quirk, 2006). Moreover, they both regulate cued reward-seeking behaviors (Ambroggi et al., 2008; Burgos-Robles et al., 2013; Ishikawa, Ambroggi, Nicola, & Fields, 2008; Peters et al., 2009).

If the roles of the amygdala and mPFC overlap so much, what is the computational advantage of having both? One possibility is that mPFC and amygdala neurons code information differently, allowing them to implement distinct sets of input-output functions.

While many emotional conditioning studies have examined the coding properties of amygdala and, to a lesser extent, mPFC neurons, most relied on simple conditioning tasks that featured one or just a few stimuli and behaviors, giving the misleading impression that neurons encode single, easily interpretable dimensions (Pare & Quirk,

2017). By contrast, studies relying on more complex or diverse experiment paradigms, revealed that amygdala (Grundemann et al., 2019; Kyriazi, Headley, & Pare, 2018; A. Saez, Rigotti, Ostojic, Fusi, & Salzman, 2015) and mPFC neurons (Ma, Hyman, Durstewitz, Phillips, & Seamans, 2016; Ma, Hyman, Lindsay, Phillips, & Seamans, 2014; Rigotti et al., 2013) encode various conjunctions of task features or behaviors, in effect implementing high-dimensional representations.

Unfortunately, other than simple conditioning studies, the coding properties of mPFC and amygdala neurons have not been examined in the same task yet. A few studies compared the coding properties of amygdala and anterior cingulate or orbitofrontal neurons in non-human primates (Klavir, Genud-Gabai, & Paz, 2013; Livneh, Resnik, Shohat, & Paz, 2012; Munuera, Rigotti, & Salzman, 2018; Pryluk, Kfir, Gelbard-Sagiv, Fried, & Paz, 2019; R. A. Saez, Saez, Paton, Lau, & Salzman, 2017), revealing significant differences between them. For instance, whereas in the amygdala, the same neurons encode social hierarchical rank and the reward value of non-social stimuli, in the orbitofrontal and anterior cingulate cortex, they only encode reward value (Munuera et al., 2018).

In this thesis, I use the RRI task to improve our understanding of how conditioned emotional stimuli and behaviors are processed by the amygdala and mPFC. In Chapter 3, I examine how LA and BL neurons encode conditioned stimuli and behaviors in the RRI task. Given that mPFC is also a critical node in regulating emotional memories, in Chapter 4 I compare the coding properties of mPFC and amygdala neurons using the RRI task. Finally, in the Discussion I integrate my findings from these two structures and discuss how they work together and independently to mediate emotional behaviors.

Chapter II:

General methods

2.1 Subjects

Procedures were approved by the Institutional Animal Care and Use Committee of Rutgers University, in compliance with the Guide for the Care and Use of Laboratory Animals. Ten adult male (300 – 350 grams) and two adult female (280 – 320 grams) Long Evans rats (Charles River Laboratories) were housed individually with *ad libitum* access to food and water and maintained on a 12 h light/dark cycle. Animals were habituated to the animal facility for one week and then to handling 15 minutes daily for 3-5 days. One week prior to training in the RRI task, animals were placed on water restriction while maintaining them at $\geq 85\%$ of initial body weight. The water restriction protocol consisted of 6 consecutive days of restriction followed by one day of *ad libitum* access to water. All experiments were performed during the light cycle.

2.2 Risk Reward Interaction Task

The apparatus (**Fig. 3**) consisted of a dimly lit rectangular arena (90 cm in length x 30 cm in width) with high walls (60 cm), a floor of metal bars (5 mm od, spaced 8 mm) and no ceiling. The floor was divided into 3 equal sectors (30 cm x 30 cm), with an array of light-emitting diodes (LEDs) below each sector (**Fig. 3A**). At each end of the arena,

there was a water port with an LED behind it (**Fig. 3A**). A programmable microcontroller (Arduino, SparkFun, Niwot, CO) controlled the activation of the LEDs, the delivery of water through the water ports (60 μ l), and the delivery of footshocks in each sector independently (0.4 mA, 10 s).

Reward availability was signaled by activation of an LED (CS-R) behind the active water port for 20 seconds. The reward was delivered 10 seconds after CS-R onset through a dipper which retracted at the end of the CS-R. Similarly, shock delivery was signaled by an LED (CS-S) below one of the floor sectors for 20 seconds. A mild footshock was delivered to the corresponding sector 10 seconds after CS-S onset. The shock and CS-S co-terminated. If rats were located in the sector when the CS-S activated, they could actively avoid the shock by leaving that sector in the first 10 seconds following CS-S onset. If rats were not in the shock sector at CS-S onset, they could passively avoid the shock. An overhead video camera (29.97 or 40 frames/s) recorded the behavior. A white noise maker was used to mask any sounds from the room during the task.

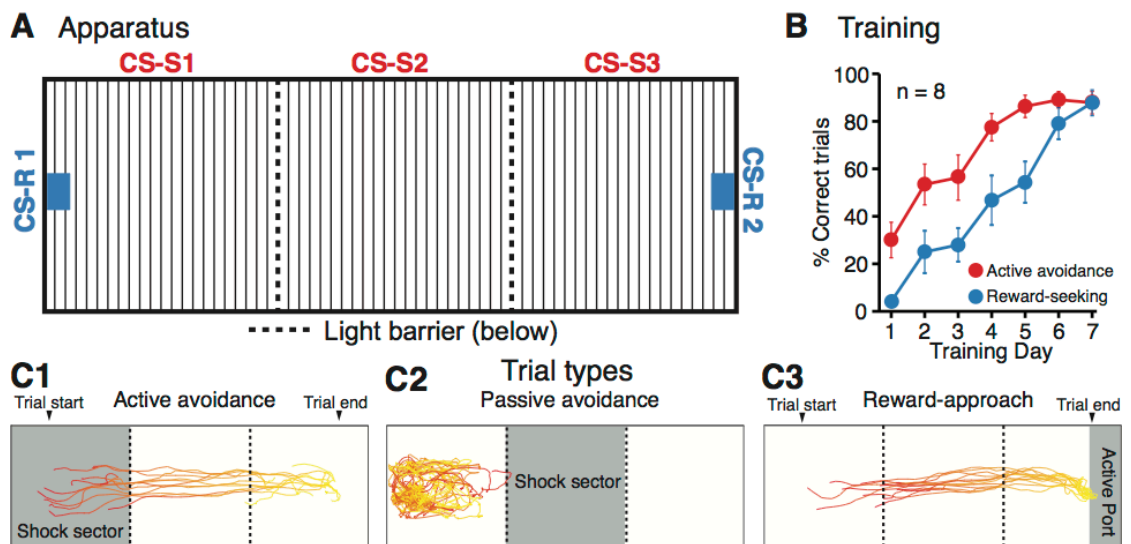


Figure 3. Risk-reward interaction task.

(A) Apparatus. LEDs at different locations signal reward availability (blue; behind left wall, CS-R1; behind right wall, CS-R2) or an impending shock (red; under the three different floor sectors, CS-S1-3). (B) Acquisition of aversive (red) and appetitive (blue) CRs. Percent correct trials \pm SEM (y-axis) as a function of training day (x-axis). (C) Examples of CRs. C1, active avoidance. C2, passive avoidance. C3, reward approach. Superimposed trajectory (start, red; end, yellow) of a rat on multiple types of trials where the rat's starting position was in sector 1, on the left.

The day before training began, rats were habituated to the arena for 1 hour. No light stimuli, water rewards, or shocks were delivered during habituation. The white noise maker was on and the lights were off for the duration of habituation. Behavioral training on the RRI task occurred over seven days. During this period, rats gradually learned to retrieve water rewards at each of the two ports and avoid shocks from each of the three sectors when signaled by the corresponding LED. The first two days, rats were trained for two 15 minute sessions (9 shock trials, 6 reward trials); one in the morning and one in the afternoon. Once animals acquired the active avoidance behavior, the sessions were extended to one hour (48 shock trials, 32 reward trials). Training continued on days 3 through 7 with 1 one-hour long sessions per day until animals reached at least 85% performance, on all trial types combined, usually within 7 days (**Fig. 3B**). For learning curves of animals used in chapter IV, see **figure 16B**.

2.3 Surgical procedures

Once they reached criterion, animals were implanted with silicon probes aimed at the BLA, mPFC, or both. Rats were anesthetized with a mixture of isoflurane and O₂ and administered atropine sulfate (0.05mg/kg, i.m.) to aid breathing. They were then placed in a stereotaxic apparatus with nonpuncture ear bars. A local anesthetic (bupivacaine,

S.C.) was injected in the scalp. Fifteen minutes later, an incision was made to expose the skull and a craniotomy performed over the mPFC and/or amygdala. One of three types of silicon probes (see below) was attached to a 3D-printed microdrive, and aimed just dorsal to the BLA (coordinates in mm relative to bregma: AP -2.2 to -3.6, ML 5.2, DV 6.0) or mPFC (coordinates: AP +2.3 to +3.7, ML 0.6 10° angle, DV 3). The microdrive allowed us to lower the probe(s) gradually between recording sessions. The three types of silicon probes we used are Buzsaki32L (**Fig. 4A**), Buzsaki64L (**Fig. 4B**), and a custom-designed probe (**Fig. 4C**). The Buzsaki32L and 64L probes respectively consist of four and eight shanks spaced 200 μm apart. The custom-designed probe has 4 shanks with 16 channels on each shank arranged in a tetrode formation. Each tetrode is 333 μm from the neighboring tetrode and the inter-shank distance is 250 μm . Rats are allowed 1-2 weeks to recover from surgery.

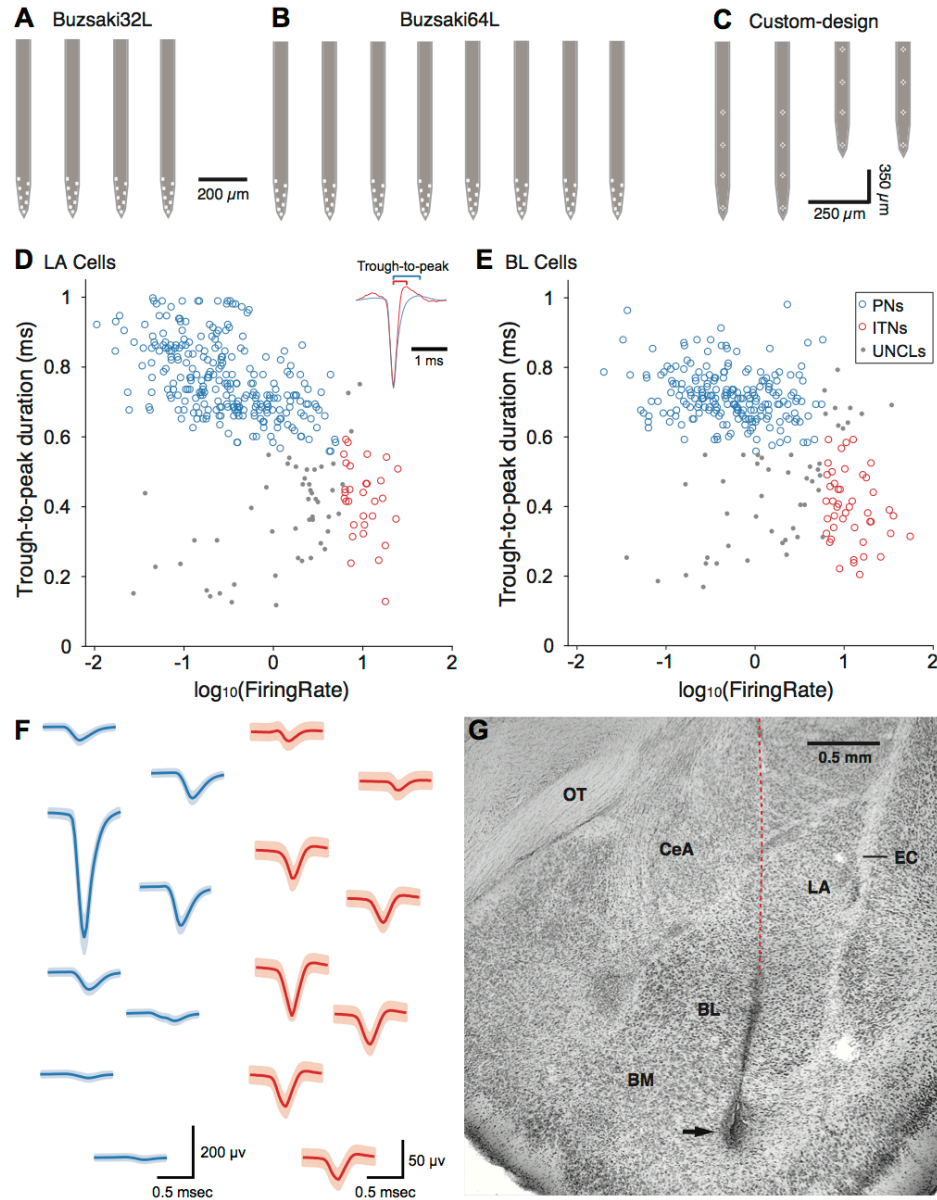


Figure 4. Methods used to record, classify, cluster, and locate BLA neurons.

The same methods are used for mPFC cells. (A-C) Three different types of silicon probes. (D,E) Classification of LA (D) and BL (E) cells in presumed projection cells (blue) or fast-spiking interneurons (red) based on spike duration (y-axis; through to peak interval; see inset; cut-off, 0.55 ms) and firing rate (x-axis; cut-off, 6 Hz). Cells that do not satisfy both criteria (unclassified, gray) are not considered further. (F) Average spike waveforms \pm SEM of a clustered projection cell (blue) or interneuron (red), as simultaneously recorded on eight neighboring leads of the same silicon probe shank. (G) Coronal section of the amygdala showing trajectory (dashed line) and final position (arrow marking electrolytic lesion made at the end of the experiment) of a silicon probe shank in the BLA. Abbreviations: BL, basolateral nucleus; BM, basomedial nucleus; CeA, central nucleus of the amygdala; EC, external capsule; OT, optic tract.

2.4 Unit recordings and clustering

Once animals recovered from the surgery, unit recordings begin. Red and green LEDs on the rat's headcap during the recording sessions allowed us to track the rat's activity throughout the task. Red LEDs were always positioned on the rostral end of the headcap and green LEDs on the caudal end. This allowed us to track head position by taking the midpoint between the red and green LEDs, and head direction by calculating the angle between the midpoint and the red LEDs with respect to the apparatus. The silicon probe was lowered ≥ 140 μm after every recording session to avoid recording the same cells across days.

Electrophysiological data was recorded using Intan Technologies equipment (RHD2000-Series amplifier evaluation system; intantech.com). Data were sampled at 30 kHz and stored on a hard drive for offline processing. A high pass filter (300 Hz) was initially applied to the recording data, followed by a median filter (median amplitude subtracted from all channels at each sample point). One of two clustering methods were used: principal component analysis with Klustakwik, or Kilosort (<https://github.com/cortex-lab/KiloSort>). For Klustakwik, principal component analysis was used on the spikes and the first three components were clustered (<http://klustakwik.sourceforge.net/>). For Kilosort, spike waveforms were identified by matching them to template waveforms as described in Pachitariu, Steinmetz, Kadir, Carandini, and Harris (2016). For both KlustaKwik and Kilosort methods, spike clusters were refined manually using Klusters (Hazan, Zugaro, & Buzsaki, 2006). For sessions clustered using Kilosort, a custom code was written to convert Kilosort output files to Klusters-compatible files to allow manual refinement.

In order to reliably separate clusters, the autocorrelograms and cross-correlograms of each cluster were inspected. Autocorrelograms had to show a refractory period of at least 2 ms. Cross-correlograms could not have a refractory period as this indicated that the same unit was shared between clusters. If a unit had unstable spike shapes, it was excluded.

Spike duration was calculated for a given unit by selecting the channel with the largest action potentials and peak-to-trough amplitude. The time between spike trough and peak represents the spike's duration (Bartho et al., 2004). BLA units were classified as presumed projection cells or interneurons based on their baseline firing rates (6 Hz cutoff) and trough-to-peak duration (0.55 ms). For PL units, a firing rate cutoff of 10 Hz and 0.5 ms trough-to-peak interval were used. Striatal units were divided into fast-spiking interneurons and presumed medium spiny neurons based on interspike interval and trough-to-peak durations. Based on Berke (2008), highly active cells with less than 2% of their interspike intervals longer than 1s were classified as interneurons. The trough-to-peak duration used for striatal units is the same as for BLA cells (0.55 ms). All cells that failed to meet both criteria were labeled unclassified and excluded from analyses.

In a prior study (Amir, Headley, Lee, Haufler, & Pare, 2018), we tested the reliability of these classification criteria on a different sample of rat BL neurons recorded with the same methods. Using cross-correlograms in ~17,500 pairs of BL neurons, we looked for evidence of monosynaptic inhibition from putative principal cells to other cells (a case where an interneuron would be misclassified as a projection cell) or excitation from putative interneurons to other cells (indicating a principal cell misclassified as an

interneuron). Only 0.99% of principal cells and 3.6% of interneurons with putative connections were found to be misclassified.

With respect to the usefulness of spike duration to distinguish principal BL neurons, only one study (Bienvenu, Busti, Magill, Ferraguti, & Capogna, 2012) directly compared the duration of extracellularly measured spike waveforms in principal cells (n=23) and different types of interneurons (n=51) of the basolateral amygdala. On average, all four classes of BLA interneurons examined in this study (pavalbumin positive, calbindin positive, axo-axonic, amygdalo-striatal projecting) generated spikes of shorter duration than principal cells. Of the 51 interneurons described in this study, only two interneurons had a spike duration as high as the principal cell with the *lowest* spike duration. Together, these earlier findings support the reliability of the criteria we used to distinguish principal cells and interneurons.

2.5 Histology

At the end of experiments, while rats were anesthetized, electrolytic marking lesions were made on either the most dorsal or ventral electrodes, alternating on each shank (10 μ A for 16 s), so that lesions marking different shanks were distinct. Twenty-four to 48 hours after lesions, rats were perfused-fixed through the heart, their brains extracted and cut on a freezing microtome. The sections (80 μ m) were counterstained with 1% thionin solution. Only neurons histologically determined to have been recorded in the structures of interest were analyzed.

2.6 Behavioral Analyses

All behaviors were analyzed using a custom made graphical user interface (GUI) in MATLAB (The MathWorks, Inc., Natick, Massachusetts, U.S.A). Behavior start and end times were determined in the GUI by indicating the first and last frame of the video recording when the rat initiated and ended a behavior. A total of five behaviors were considered for analyses: freezing (Frz), active avoidance (AA), passive avoidance (PA), reward approach (RA), and reward anticipation (RAnt). Behavior start and end times were defined as follows: for AA, the start was noted as the first frame when the rat began to move off the lit sector, and the end of the behavior was denoted as the first frame when the rat ended the continuous avoidance behavior on one of the unlit sectors. PA start time was denoted as the first frame of a CS-S trial where the animal was away from the lit sector and ended either at the end of the trial (provided the animal remained away from the lit sector for the entire trial duration) or until the animal approached or engaged with the lit sector. Frz was denoted as any period of immobility except for respiration lasting at least one second during CS-S trials. For RA, the start time was the first frame the rat began movement toward the reward port and the end was the frame when his/her nose reached the port. RAnt began when the rat placed both front paws on the reward port and waited there until reward delivery.

2.7 Classification of valence neurons

A cell was defined as a R-cell if, relative to the 5 s immediately preceding the CS-R, it had a significant increase in firing rate (based on rank-sum test $p < 0.005$) either during the first 1s after CS-R onset or during the initial 10s CS-R period and an inhibition

or no change to the CS-Ss. S-cells were similarly defined using the opposite valence CSs. AA-cells and RA-cells were defined as cells with a significant increase in firing rate for 1s after behavior onset compared to 2s prior to behavior onset.

2.8 Generalized Linear Model (GLM)

A regularized regression, group Lasso, with Poisson distribution (grpreg R package; Breheny & Huang, 2015) and ten-fold cross validation was used to fit the spiking of individual cells for the duration of the task. Spiking was binned (50 ms bins) across the entire recording session and stimuli and behaviors were indicated with ones when they occurred. After Park, Meister, Huk, and Pillow (2014), the stimulus and behavior events were convolved with basis functions defined by log-time scaled raised cosine bumps separated by $\pi/2$ radians (50 ms). Each event kernel was represented as a linear combination of basis functions spanning a duration of time (see below)

The model was fit by minimizing the value β in

$$Q(\beta) = \frac{1}{2n} ||y - X\beta||^2 + \lambda \sum_j \sqrt{K_j} ||\beta_j||$$

where y represents the spike train across time, X is the design matrix with the basis functions for stimuli and behaviors, β s are the least-squares regression coefficients, λ is the regularization penalty for an L_1 norm, n is the total number of predictors, and $\sqrt{K_j}$ is a matrix of the kernels for each predictor j and it serves as a way to normalize across groups of different sizes. The lasso penalty parameter (λ) to the Euclidean (L_1) norm was chosen based on the lowest cross-validation error and it was applied to each group, creating sparsity and variable selection at the group level (Breheny & Huang, 2015;

Robert Tibshirani, 1996; Ming Yuan & Yi Lin, 2006). Furthermore, the penalty is applied to the entire group such that if group j is selected then all β s are either set to zero or all β s are non-zero. Cross-validation sets were assigned by dividing the recording session into ten equal segments.

For BLA data in chapter III, there were a total of 17 parameters used in the GLM including speed, position, CS-R1, CS-R2, CS-S1, CS-S2, CS-S3, US-R1, US-R2, US-S1, US-S2, US-S3, AA, PA, Frz, RA, and RAnt. For chapter IV data, which included the original BLA dataset and all newly recorded PL and BLA units, there were a total of 22 grouped parameters used in the GLM: speed, position, CS-R1, CS-R2, CS-S1, CS-S2, CS-S3, US-R1, US-R2, US-S1, US-S2, US-S3, AA, Frz, RA, RAnt, CS-R2 RA, CS-R2 RAnt, CS-S2 AA, CS-S2 Frz, CS-S3 AA, CS-S3 Frz. In chapter IV, the aim was to capture interaction terms in the model hence, the original dataset of chapter III was re-fit with a new Lasso GLM. Interaction terms were computed by multiplying the stimulus variable with the corresponding behavior variable (i.e. $\text{CS-S1} * \text{AA} = \text{CS-S1 AA}$). In order to avoid perfect collinearity with interaction terms, we did not include CS-R1 and CS-S1 variables interaction terms.

GLM Basis Functions and Kernels. The GLM kernels for each stimulus and behavior type that best fit spiking activity were created by combining a set of pre-determined basis functions. We used two different sets of basis functions to represent CSs and CRs. Stimulus basis functions covered the initial 10 s, have a sharp onset and narrow width which becomes smoother and wider across time (**Fig. 5A**), reflecting stimulus responses that tend to have sharp onsets and decay slowly. Behavior basis functions

extended before and during the behavior onset to capture spiking activity related to planning and executing the behavior (**Fig. 5B**). However, they were bounded by the start of a trial. These functions were then scaled by the model's beta values and summed within time bins to create a single kernel for each stimulus and behavior (see **Fig. 5C-D** for example stimulus and behavior kernels from a cell fit by the model). Most importantly, if a cell did not encode a parameter, the group Lasso GLM gives beta values of 0 for the corresponding basis functions. Thus, only parameters that best fit observed spiking of a given cell were selected by the model, allowing us to perform dimensionality reduction.

To compare the model fit to observed spiking, we used the coefficient of determination (R^2). Kernels fit with all model parameters for appetitive and aversive trials during the 10s CS period were adjusted for firing rate and concatenated. Similarly, observed PETHs for the 10s CS period during all appetitive and aversive trials were concatenated and smoothed with a Gaussian window of 550 ms. These model-estimated kernels and observed PETHs were used to calculate the coefficient of determination as follows:

$$R^2 = 1 - \frac{\sum (y_i - f_i)^2}{\sum (y_i - \bar{y})^2}$$

where y_i represents the observed PETH, f_i is the model-estimated kernel, at different time points, i , and \bar{y} is the mean of the observed PETH.

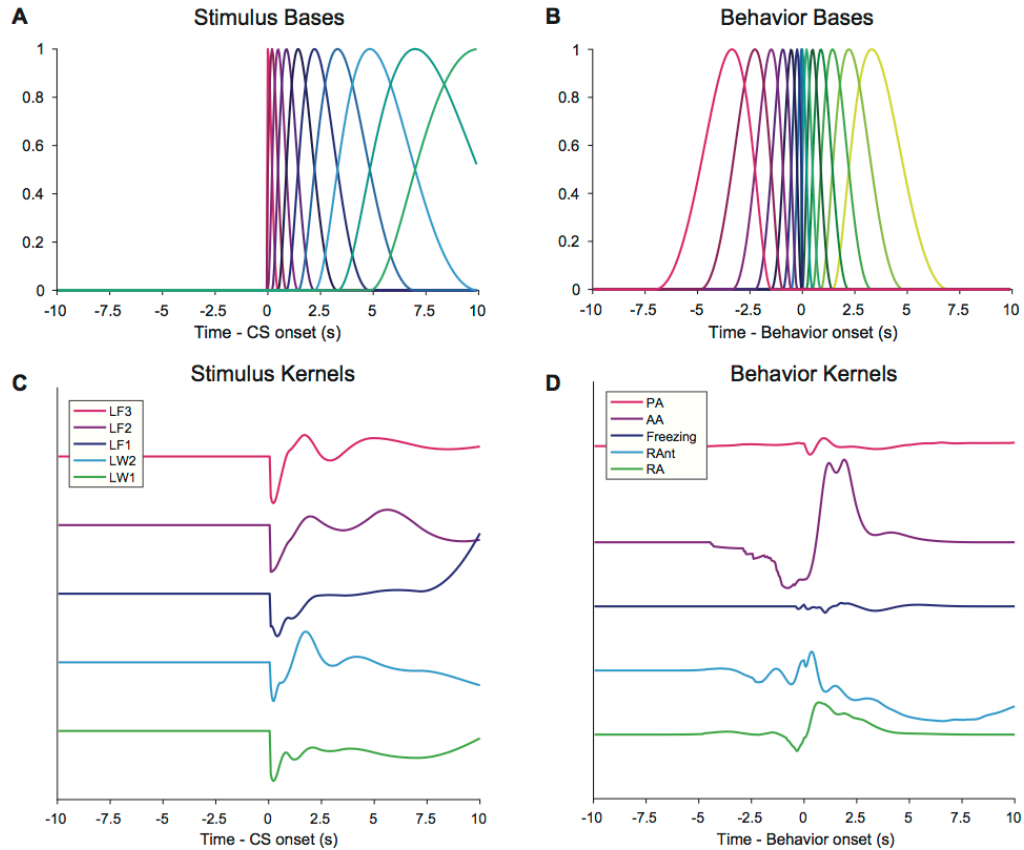


Figure 5. GLM basis and kernel functions.

(A) The 10 basis functions used to fit stimulus responses starting at the onset of the stimulus presentation and extending out to 10 seconds. (B) The 14 basis functions used to fit neural responses related to behaviors; expanding before behavior onset to capture preparatory signals. (C) Example kernels for each stimulus produced by multiplying the stimulus basis functions and beta values computed by the GLM. (D) Example behavior kernels fit by the GLM for the same cell as in C. Note, the kernels capture the transient inhibition prior to active behavior onset (active avoidance, purple trace; reward approach, green trace) and excitation during the execution of the behaviors.

2.9 Multidimensional scaling (MDS)

To visualize and quantify the coding of task dimensions at the population level for each region in chapter III, we mapped our codes (see **Fig. 14A4**) into a low dimensional space created using multidimensional scaling (MDS). This technique projects a set of points in a high dimensional space into a low dimensional space, while trying to preserve their pairwise distances. For a given unit we constructed a feature vector by taking the

GLM-estimated peak modulations by the five stimuli (CS-R1, CS-R2, CS-S1, CS-S2, CS-S3) and five behaviors (RA, Rant, AA, Frz, speed) and ranked each dimension relative to the other units in the data set. We then calculated the Euclidean distance between all pairs of feature vectors within a region. This set of distances was passed to the MDS algorithm (non-metric, squared stress criterion) to reduce the distribution from ten to two dimensions, while still preserving the relative distances between the cells. Then, for each unit type and region we added a third dimension that corresponded to one of the codes suggested by the correlation matrices (behavior valence, stimulus valence, stimulus/behavior) or previous studies (speed, active vs. passive behaviors; Amir, Lee, Headley, Herzallah, & Pare, 2015). This created a collection of points in a three-dimensional space, and we fit a plane (code-value = $\beta_1\text{MDS-Dim1} + \beta_2\text{MDS-Dim2} + \text{intercept}$) to those points to determine if the feature of interest was systematically mapped across the population (Borg & Groenen, 2005). The magnitude of the slope of the plane, $\sqrt{\beta_1^2 + \beta_2^2}$, indicated the degree to which that feature was mapped across the population. By comparing the difference in the direction, $\arg(\beta_1 + i\beta_2)$, between two feature planes, we could determine the degree of overlap in their representations at the population level.

2.10 Normalized peak firing rate modulations.

The absolute peak of each CS and CR kernel is normalized by the baseline firing rate as follows: (Peak-Baseline)/Baseline. The sign of the modulation is preserved in the normalized peak in order to identify cells excited or inhibited in relation to each task feature. Normalized peak modulations ≤ 0.001 are considered non-significant and are set

to zero. The average modulations to the task features are computed by taking the mean of the absolute value of all significant peak modulations for each parameter. For **figure 12**, the *observed* normalized peak firing rate during active avoidance and reward approach are plotted on the x and y axes, respectively. These values are calculated by binning the observed spiking in 50ms bins and referencing PETHs to behavior onset. The mean firing rate during the behavior is normalized by the firing rate during the 5 s preceding behavior onset, similarly to the normalized peaks for the GLM.

2.11 Similarity Matrices

We computed similarity matrices using a Spearman correlation of the peak firing rate modulations by all stimuli and behaviors across cells. In chapter III, the p-value is corrected for multiple (45) comparisons ($p=0.001$). In chapter IV, no p-values were calculated for the similarity matrices.

2.12 Number of Features Encoded.

To identify how many features cells encoded in chapter III, we used the significant normalized peak firing rate modulations of each cell task features: CS-R, CS-S, RA, RAnt, Frz, AA, and PA. For the CSs, we averaged modulations within CS types (CS-R1-2; CS-S1-3). The number of significant excitatory modulations was counted for each cell and that value represented the number of features encoded. To test if the distributions of features encoded differs between structures, we used the Kullback-Leibler divergence as a test statistic. A null distribution was created by shuffling the location of each cell randomly 10,000 times and computing a divergence score after each

permutation. This gave us a distribution of divergence scores which we then used to compare the observed scores. If the observed score was outside the two-sided 95th percentile of the null distribution, it was considered significant.

2.13 Gini and dimensionality indices

Gini Index. To understand how the code was distributed across the population we used the Gini index. This index quantifies the coding distribution across the neuronal population and ranges from 0 to 1, with 0 representing a perfectly equally distributed code among neurons, and 1 representing a perfect inequality of code distribution. The Gini index is calculated using the following function:

$$Gini\ index = \left(n + 1 - 2 * \frac{\sum_{i=1}^n (n + 1 - i) * Y_i}{\sum_{i=1}^n Y_i} \right) * \frac{1}{n - 1}$$

where Y_i represents the peak firing rate modulation for each task feature n . The Gini index was computed for each neuron individually based on the GLM firing rate modulation to each task variable. The distribution of the Gini indices across all cells in each structure is plotted in **Fig. 24B** (top). A rank-sum test was used to test whether the Gini index distributions differed across brain regions ($p < 0.001$).

Dimensionality Index. To identify how many features cells encoded in chapter IV, we used the significant normalized peak firing rate modulations of each cell for all task features: CS-R1, CS-R2, CS-R1 RA, CS-R1 RAnt, CS-R2 RA, CS-R2 RAnt, CS-S1, CS-S3, CS-S1 Frz, CS-S1 AA, CS-S3 Frz, CS-S3 AA. The number of non-zero modulations were counted for each cell and that value represented the number of features encoded.

The dimensionality index was then plotted in **Fig. 24B** (bottom) as the distribution of the number of features encoded across all cells.

2.14 D-prime analyses

The d-prime (d') metric was used to measure the separation between different coding dimensions based on the distributions of the correlation coefficients in the similarity matrices (Keene et al., 2016; McKenzie et al., 2014). The coding dimensions tested were *valence* (chapter III: CS-R1, CS-R2, RA, RAnt vs. CS-S1, CS-S2, CS-S3, Frz, AA; chapter IV: CS-R1, CS-R2, CS-R1 RA, CS-R1 RAnt, CS-R2 RA, CS-R2 RAnt vs. CS-S1, CS-S3, CS-S1 Frz, CS-S1 AA, CS-S3 Frz, CS-S3 AA), *valence-CS* (chapter III: CS-R1, CS-R2 vs. CS-S1, CS-S2, CS-S3), *valence-behavior* (chapter III: RA, RAnt vs. Frz, AA), and *active vs. passive behaviors* (chapter III: RA, AA vs. RAnt, Frz; chapter IV: CS-R1 RA, CS-R2 RA, CS-S1 AA, CS-S3 AA vs. CS-R1 RAnt, CS-R2 RAnt, CS-S1 Frz, CS-S3 Frz). The d' metric was calculated for each coding dimension by comparing the degree to which within-coding dimension correlations exceeded those between-coding dimension correlations, as follows:

$$d' = \frac{\mu_{within} - \mu_{between}}{\sqrt{\frac{1}{2}(\sigma_{within}^2 + \sigma_{between}^2)}}$$

Where μ_{within} and $\mu_{between}$ are the means of the correlation coefficients for the within and between coding dimension features, respectively, and their corresponding variances, σ_{within}^2 and $\sigma_{between}^2$. To identify d' values significantly different from 0, we compared the observed d' values to a bootstrap distribution in which normalized peak values are randomly sampled 10,000 times, and similarity matrices and d' values were then re-

computed for each bootstrap sample. The p-value of the observed d' was calculated based on the bootstrapped distribution and considered to be significant if $p < 0.05$. To compare if two d' values were significantly different from each other, we computed a shuffled distribution of d' -difference values (10,000 permutations). If the observed d' -differences were $>95\%$ of the shuffled distribution then the d' -differences were considered significant.

2.15 Principal Component Analysis (PCA)

To visualize the coding of task dimensions at the population level for each region, we mapped the firing rate into a low dimensional space (3-dimensions) created using principal component analysis (PCA) across the trial duration (Rozeske et al., 2018; Zhang & Li, 2018). Spiking was extracted from 5 s before to 10 s after CS-onset. The firing rates for baseline, CS-onset, and behaviors (AA and RA) were normalized by binning the firing rate into 20 bins for each time epoch. This time normalization was conducted to ensure that spiking to behaviors was always aligned despite trial-to-trial variations in the timing of behaviors. The binned spikes were then z-scored using the 5 s baseline period and averaged across trials. A boxcar function convolution was applied to the z-scored spiking before conducting the PCA analysis. The first 3 dimensions of the PCA scores were smoothed with a Gaussian window of 10 bins for visualization. All further calculations using the PCA scores were applied to the raw, unsmoothed scores.

Euclidean Distance Based on PCA Scores. The 3-dimensional Euclidean distance was calculated at every time bin between each trial type (CS-R1, CS-R2, CS-S1, and CS-S3) using the raw, unsmoothed PCA scores. The distances across time were then

averaged to calculate a mean distance value for each trial type comparison, which is plotted in **Fig. 26B**.

2.16 Support Vector Machine Decoder

Support vector machine analyses were conducted in a time-normalized fashion similar to PCA, but instead of creating 20 bins for each time epoch, 1 bin was used, resulting in 3 bins total (baseline period, CS-onset period, active behavior period (AA or RA)). The SVM decoder was conducted similarly to Meyers, Freedman, Kreiman, Miller, and Poggio (2008). Briefly, a pseudo-population of 150 cells with replacement was selected from each brain region. For each cell, 5 trials of each type were selected (5 CS-R1, 5 CS-R2, 5 CS-S1, and 5 CS-S3 trials). For the trial type comparison, the two appetitive trial types were compared to each other, and the two aversive trial types were compared to each other. This process was repeated 50 times. The appetitive and aversive decoder performance repetitions were combined to create the trial identity category. The mean decoder accuracy and SEM were computed using the 100 total repetitions. For the valence category, the CS-R responses were compared to the CS-S responses using 10 trials of each type.

To test for significantly different decoding accuracies between brain regions, we bootstrap resampled the decoder accuracy repetitions of two brain regions 10,000 times, and each time calculated a difference score. If the 95th percentile of this null distribution crossed zero then the difference was not considered significant, if it did not cross zero then it was considered significant.

Population size decoder. The size of the pseudo-population was varied from 1, 5, and then 10 to 150 cells in increments of 10 cells. Decoder performance was calculated with each pseudo-population size and the mean and SEM were calculated from the 50 repetitions.

2.17 Statistical Analyses

Group data are reported as average \pm SEM. All neurons with stable firing rates and spike shapes were included. Firing rates reported logarithmically were natural logarithms. All statistical tests were two sided. Different procedures were used to assess statistical significance depending on the type of data, as specified below.

Observed behavior- or CS-related changes in firing rates. To determine if individual neurons showed significant behavior- or CS-related changes in firing rates, we used the Wilcoxon rank-sum test ($p < 0.005$) on the observed binned spiking locked to either the behavior or CS onset. Spiking was binned in 50 ms windows. A baseline of 5 s before the event onset was used to compare binned spiking with either 1 s after CS/behavior onset or 10 s after CS onset.

Incidence of valence cells. In order to determine if the proportion of valence cells was significantly higher than expected by chance given the proportions of cells whose activity modulated the various CSs and CRs, we permuted the model-estimated peak firing rate modulations for all task variables across cells 10,000 times. This shuffling procedure randomized any potential associations between encoded variables and yielded a null distribution for the incidence of valence-coding cells. Using this null distribution,

we then calculated the percentile of the observed proportions. If this value fell outside the two-sided 95th percentile of the null distribution it was considered significant.

Assessment of population response maps. We estimated the distribution of the slopes using bootstrap resampling. The plane was refit 500 times using different collections of neurons that were resampled with replacement. Each of the coding planes were recalculated for the same collection of resampled data points, generating matched pairs. For each code, these were used to generate 95% confidence intervals. To determine if a coding dimension was significantly mapped across the MDS space, we permuted code values across neurons and recalculated the planes (500 times), yielding the distribution of null planes. To determine if two codes were significantly different in their strength or direction, we took the difference in the magnitude or angle (using circular distance) of the slopes between the matched pairs and found the probability of a difference of zero. When comparing between codes, a Bonferroni correction for multiple comparisons (10 potential comparisons) was performed on all p-values.

Chapter III:

Multidimensional coding by basolateral amygdala neurons

3.1 Rationale

The basolateral complex of the amygdala (BLA) is a structure involved in the acquisition and expression of conditioned responses (CRs) to stimuli (CSs) that predict aversive or appetitive outcomes. These CRs are thought to result from the activation of specific subsets of valence-coding BLA neurons. Under this model, the responses of BLA cells to CSs and the activity that drives CRs are closely related. In this chapter, I test this hypothesis using a novel paradigm which can dissociate BLA activity related to stimulus responses and conditioned behaviors. Eight rats were trained on the RRI task while recording neurons in LA, BL, and the striatum using movable silicon probes. See chapter II for a detailed description.

3.2 Rats quickly learn the Risk Reward Interaction task

In the RRI task (**Fig. 3**), light CSs indicated different outcomes based on their location. When the CS appeared below one of three shock sectors (CS-S; **Fig. 3A**, red), it signaled an impending footshock in that sector. When a CS was presented behind the left or right wall (CS-R; **Fig. 3A**, blue), it indicated that a water reward would be delivered at that location. The rats' position and head direction were monitored throughout the task.

After training, rats ($n=8$) displayed different defensive and appetitive behaviors during the RRI task. On CS-S trials, if rats were on the shock sector at trial onset, they generally avoided the shock by moving to one of the unlit sectors (**Fig. 3C1**). In 19% of such *active avoidance* trials, animals showed *behavioral freezing* prior to avoiding the shock. Alternatively, if rats were already on one of the unlit sectors at the start of the CS-S, then they typically displayed *passive avoidance*; that is, they stayed away from the lit sector until the trial ended (**Fig. 3C2**).

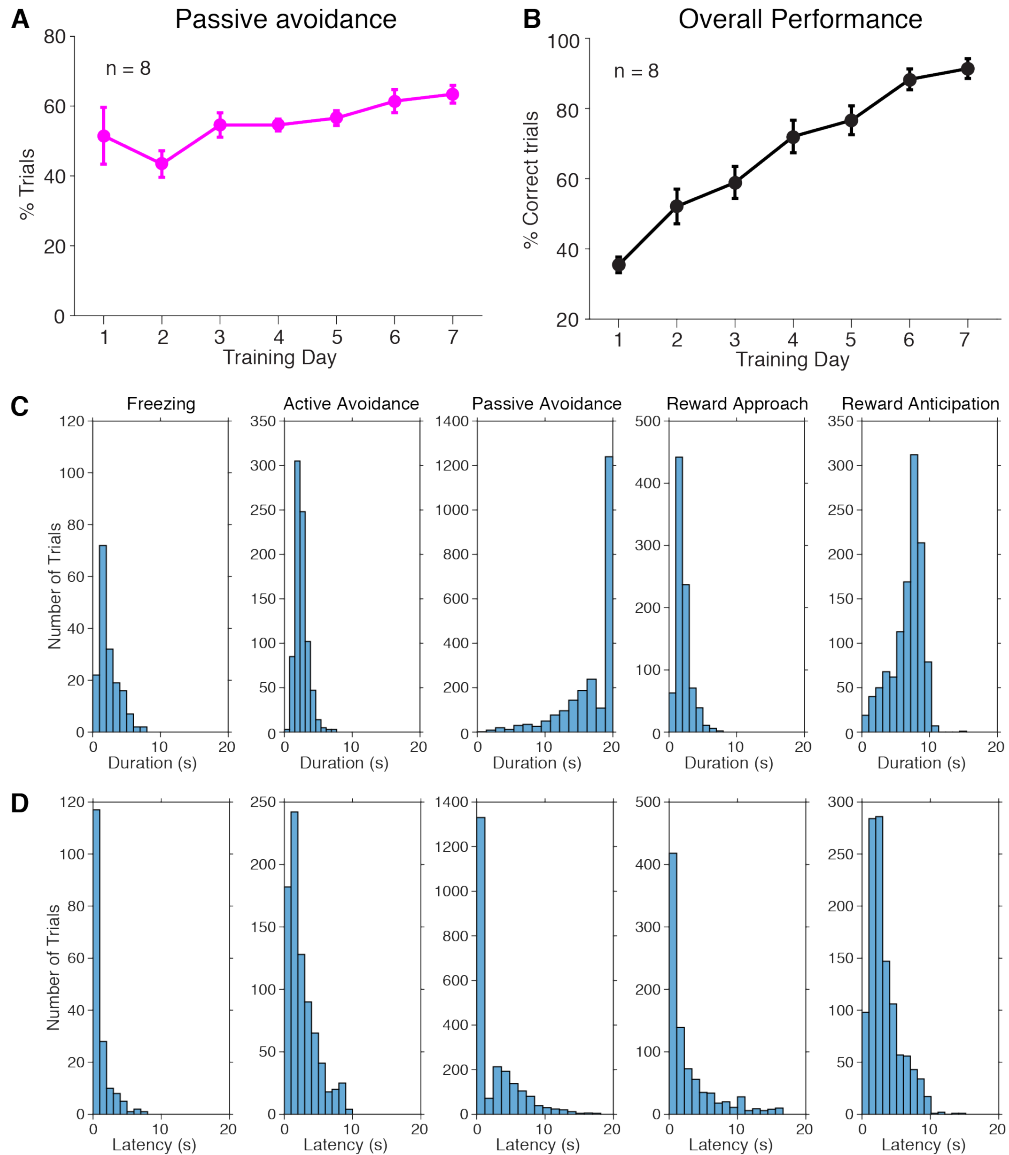


Figure 6. Incidence, duration, and latency of different types of CRs during training. (A) Incidence (% trials; y-axis) of passive avoidance as a function of training day (x-axis). (B) Overall performance (% correct trials; y-axis) as a function of training day (x-axis). Frequency distributions of the duration (C) or latency from CS onset (D) of the various CRs monitored in this study.

During appetitive trials, rats ran to the reward port shortly after CS-R onset. This *reward approach* behavior (**Fig. 3C3**) was then followed by *reward anticipation*, defined as rats placing their two front paws onto the reward port and waiting there until reward delivery.

During the training phase, rats acquired active avoidance faster than reward seeking (**Fig. 3B**; two-way ANOVA, Behavior Type x Session, $F_{(1,104)}=33.32$, $p=8.15 \times 10^{-8}$). Passive avoidance remained stable throughout training (**Fig. 6A**). As detailed in **figure 6C-D**, conditioned behaviors varied significantly in their duration (one-way ANOVA, $F_{(4,5259)}=31.74$, $p=3.62 \times 10^{-26}$) and latency from CS onset (one-way ANOVA, $F_{(4,5259)}=7238.91$, $p=0$). Once rats performed $\geq 85\%$ of the trials correctly (**Fig. 6B**), they were implanted with silicon probes aimed at the BLA (**Fig. 4A-C**).

3.3 Activity of principal BLA cells during CSs

We recorded 344 cells in LA and 305 cells in BL while rats performed the RRI task. Cells were classified as putative principal neurons (PNs; LA $n=264$; BL $n=212$) or fast-spiking interneurons (ITNs; LA $n=31$; BL $n=45$) based on spike width and mean firing rate (**Fig. 4D-E**). To assess how the proportion of cells with CS-responses in the RRI task compared to previous reports on appetitive and aversive tasks, we computed the proportion of cells significantly responsive during the first second after CS onset. We found that $\sim 20\%$ of LA and $\sim 35\%$ of BL cells significantly increased or decreased their

firing rates in response to the CS-Rs or CS-Ss (**Fig. 7A-B**, top row; rank-sum tests, $p < 0.005$), consistent with prior studies (Beyeler et al., 2016; Burgos-Robles et al., 2016; Lee et al. 2016, 2017; Namburi et al., 2015; Sangha et al., 2013; Shabel and Janak, 2009). When we expanded the test window from one to ten seconds (the CS period prior to the US), the proportion of cells responsive during the CS-Rs and CS-Ss increased significantly in both LA (**Fig. 7A-B**; chi-square tests: CS-Rs, $\chi^2 = 32.59$, $p < 0.00001$; CS-Ss, $\chi^2 = 20.48$, $p = 0.000036$) and BL (**Fig. 7A-B**; CS-Rs, $\chi^2 = 50.83$, $p < 0.00001$; CS-Ss, $\chi^2 = 16.36$, $p = 0.00028$).

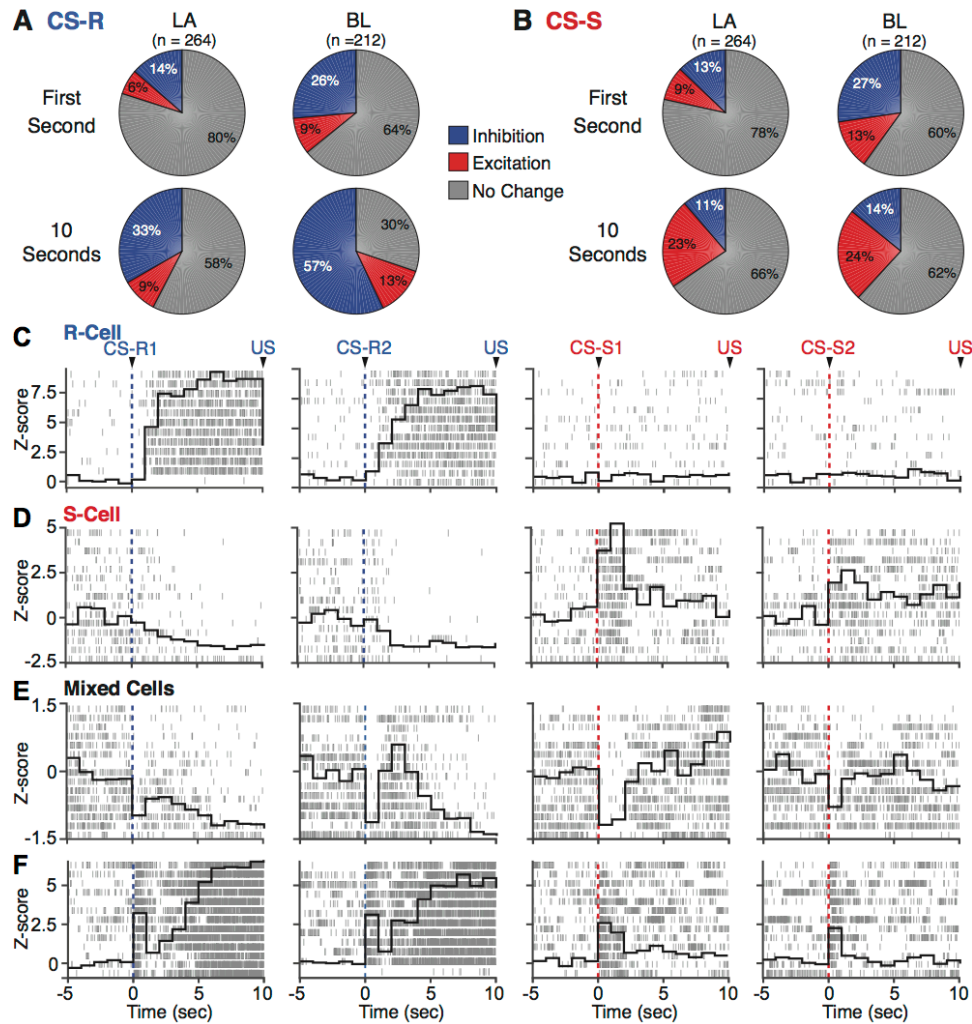


Figure 7. Activity of BLA neurons during the CS-Rs and CS-Ss.

(A) CS-R. **(B)** CS-S. Proportion of LA or BL cells with significantly altered firing rates during the first second of the CSs (**top**) or their entire duration (**bottom**). **(C-F)**, Representative examples of cells that showed significant changes in firing rates during one or more of the CSs. S-Cells **(C)** and R-Cells **(D)** are neurons whose firing rates increased selectively during the CS-Rs or CS-Ss, respectively. *Mixed cells* **(E,F)** are neurons that displayed marked trial-to-trial variations in the late part of their CS-related activity. **Ticks**, individual spike times. **Thick lines**, z-scored averages of firing rates. **Vertical dashed lines**, onset of CS-Rs (blue) or CS-Ss (red). Related to figure S2. Abbreviations: CS-R, reward-predicting conditioned stimulus; CS-S, shock-predicting conditioned stimulus; US, unconditioned stimulus.

Also consistent with prior studies, our sample included cells whose firing rates increased selectively during the CS-Rs or CS-Ss. These cells, respectively termed *R-Cells* (**Fig. 7C**; LA, 5%; BL, 10%) and *S-Cells* (**Fig. 7D**; LA, 19%; BL, 20%), showed similar changes in activity during CSs of the same valence: if a cell was responsive to one CS-R (or one CS-S), it responded similarly to the other CS-R (or CS-Ss), but displayed no response or an inhibition during CSs of the opposite valence. Furthermore, these responses were independent of head direction (three-way ANOVA for stimulus-excited cells, Head Direction x Stimulus Type x Region, $F_{(12,712)}=1.05$, $p=0.4039$) and stimulus proximity at CS onset (three-way ANOVA for stimulus-excited cells, Stimulus Proximity x Stimulus Type x Region, $F_{(2,276)}=1.4$, $p=0.2487$). However, most CS-responsive cells, hereafter termed *mixed cells* (LA, 40%; BL, 53%), displayed marked between-cells as well as trial-to-trial variations in the late part of their CS-related activity, making it difficult to classify them into separate groups. Two examples of mixed cells are depicted in **figure 7E-F**. The first shows a transient inhibition at the onset of all CS types followed by an excitation during some CS-R and CS-S trials (**Fig. 7E**). The second increased its firing rate at the onset of all CSs but displayed a late persistent excitation only during the CS-Rs (**Fig. 7F**).

3.4 Activity of principal BLA cells in relation to CRs

We reasoned that the trial-to-trial fluctuations seen in the late part of CS responses could be related to variations in the type or timing of the CRs. To test this, we computed peri-event time histograms (PETHs) of unit activity referenced to the onset of CRs instead of CSs, revealing that the activity of many PNs in LA (30.6%) and BL (31.1%) significantly increased in relation to one or more CRs relative to their firing rate immediately preceding behavior onset (rank-sum test, $p < 0.005$; see Methods). CR-activated cells were found among the three groups described above (R-cells, S-cells and mixed cells). In many of these cells (LA, 56 of 81; BL, 38 of 66), CR-locked activity occurred despite the absence of significant short-latency CS responses. Two examples of such cells are shown in **figure 8A-B**. The first cell selectively increased its firing rate at the onset of active avoidance (**Fig. 8A1**) and remained at baseline during passive avoidance trials (**Fig. 8A2**). Note the absence of increased spiking at the onset of the light stimulus (CS-S), indicated by the orange lines in **figure 8A**. The second cell did not change its firing rate during reward approach (**Fig. 8B1**), but subsequently displayed a marked and persistent activation whose onset coincided with that of reward anticipation (**Fig. 8B2**).

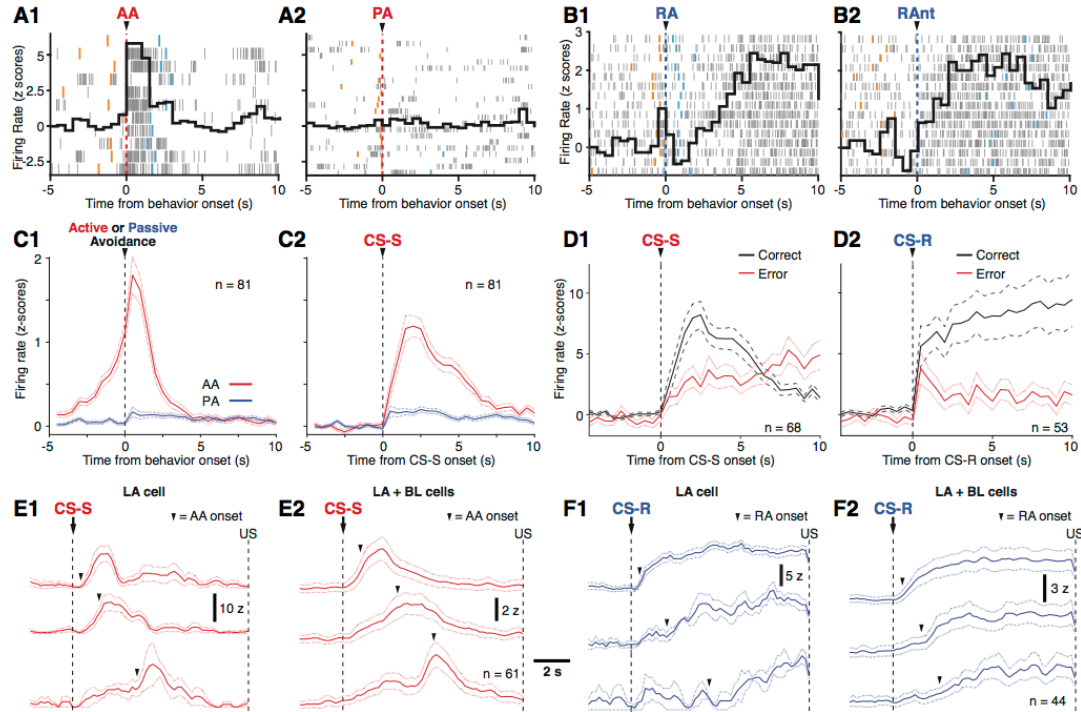


Figure 8. Behavioral correlates of unit activity.

(A,B) Individual examples of principal neurons that strongly increase their firing rates in relation to (A1) active avoidance (AA) but not passive avoidance (PA; A2) or (B2) reward anticipation (RAnt) but not reward approach (RA, B1). **Vertical dashed lines**, onset of reference behavior. **Ticks**, individual spike times (as many trials as rows of ticks are shown). **Thick lines**, z-scored averages of firing rates. **Yellow ticks**, onset of CS. **Cyan ticks**, end of reference behavior. (C) Comparison between z-scored averaged firing rate \pm SEM of principal cells ($n=81$) during AA (red) vs. PA (blue), referenced to behavior onset (C1) or CS onset (C2). (D) Comparison between z-scored averaged firing rate \pm SEM of principal cells during correct (black) and error (red) CS-S (D1; $n=68$) or CS-R (D2; $n=53$) trials. (E,F) Relation between behavior onset (arrowheads) and unit activity for CS-S (E) and CS-R trials (F). Individual principal cells are shown in E1 and F1 (z-scored average \pm SEM of multiple trials). Z-scored average \pm SEM of all available principal cells are shown in E2 and F2.

To test whether these CR-related increases in firing rates in fact represent delayed sensory responses to the light stimuli, we compared the activity of all putative CR-coding cells in three conditions: when rats emitted different CRs in response to the same CS, on correct vs. error trials, or when rats emitted the same CR at different times with respect to CS onset. For the first analysis, we took advantage of the fact that on aversive trials, rats

could avoid the US actively or passively, depending on their location with respect to the CS-S. Thus, we averaged the activity of all cells that, individually, showed significant increases in firing rates during active avoidance (hereafter termed *AA-cells* for simplicity; LA, $n=49$; BL, $n=32$) and compared their activity on aversive trials that called for an active (red, **Fig. 8C**) or passive avoidance (blue, **Fig. 8C**) response. Whether the activity of AA-cells was referenced to the onset of CRs (**Fig. 8C1**) or CS-Ss (**Fig. 8C2**), they showed significantly lower firing rates during trials that called for a passive instead of an active avoidance response (paired t-tests; CR-locked responses, $t(29)=3.373$, $p=0.002$; CS-locked responses, $t(29)=4.44$, $p=0.0001$). However, it is possible that on CS-S trials calling for an active vs. passive avoidance response, rats actually experienced a different CS. At odds with this interpretation however, we noticed that in 29% of trials where the rats were outside the shock sector, they nevertheless generated behavioral freezing in response to the CS-S, indicating that on those trials, they perceived the CS-S and interpreted it as threatening even though they were not in the shock sector.

To determine whether this contrast resulted from the rats' differing proximity to the light stimuli on active vs. passive avoidance trials, we next compared the averaged activity of all the AA-cells recorded during sessions with one or more error trials, where the rats failed to actively avoid the shock (46 LA cells, 22 BL cells, 18 sessions, 137 error trials). During the first five seconds following CS onset, by which time active avoidance responses are typically completed, AA-cells showed significantly higher firing rates on correct than error trials (**Fig. 8D1**; black, correct; red, error), even though their position relative to the light stimuli was the same in both cases (paired t-test $t(10)=8.497$, $p=6.92 \times 10^{-6}$). Furthermore, consistent with a recent report (Lee et al., 2016), this was also

true for cells with significant increases in firing rates during the CS-Rs (31 LA cells; 22 BL cells, 14 sessions, 167 error trials). In these cells, while the early sensory-related part of the response to the CS-R remained when rats failed to approach and anticipate reward delivery, the late component was nearly abolished (**Fig. 8D2**; paired t-test $t(19)=17.345$, $p=4.17 \times 10^{-13}$).

Additional evidence of the behavioral dependence of unit activity was obtained by comparing the time course of firing rates on trials where rats emitted the correct CRs at different latencies with respect to CS onset. **Figure 8E1** shows an example AA-cell whose activity increased when the avoidance behavior was emitted at different latencies across trials (black arrows). This dependence of cell activity on behavior latency was also evident when all AA-cells ($n=81$) were considered (**Fig. 8E2**). Similar results were obtained with RA-cells ($n=44$; **Fig. 8F1,2**).

3.5 Generalized linear model untangles cell activity

Even though the above suggests that the activity of BLA neurons is related to both stimuli and behaviors, the temporal overlap between CSs and associated CRs constitutes a major obstacle when determining what variables LA and BL cells actually encode with PSTHs. To circumvent these limitations, we fit the spiking of individual cells using a group Lasso generalized linear model (GLM) with ten-fold cross validation. This method is not based on PETHs, but takes full advantage of trial-to-trial variations in the type, timing and duration of the variables of interest (stimuli, evoked and spontaneous behaviors) to infer which one(s) neurons encode. It is used to fit the spiking of individual cells for the duration of the task. Importantly, this type of GLM allows for dimensionality

reduction in correlated data and encourages sparsity when identifying the task variables that are related to cell activity (Breheny & Huang, 2015; Robert Tibshirani, 1996; Ming Yuan & Yi Lin, 2006).

To test whether the GLM's coding estimates had face validity, we first compared the model's output to the results of PETHs referenced to different variables. For instance, standard analyses had revealed that the cell shown in **figure 9A** (same as in **Fig.7F**) had short-latency (behavior-independent) responses to all CSs, that it increased its firing rate in relation to reward anticipation and, more weakly, during active avoidance. The model correctly captured this complex response profile (**Fig. 9A**, CSs row 1; reward anticipation row 3, columns 4-5; active avoidance, row 3, column 1).

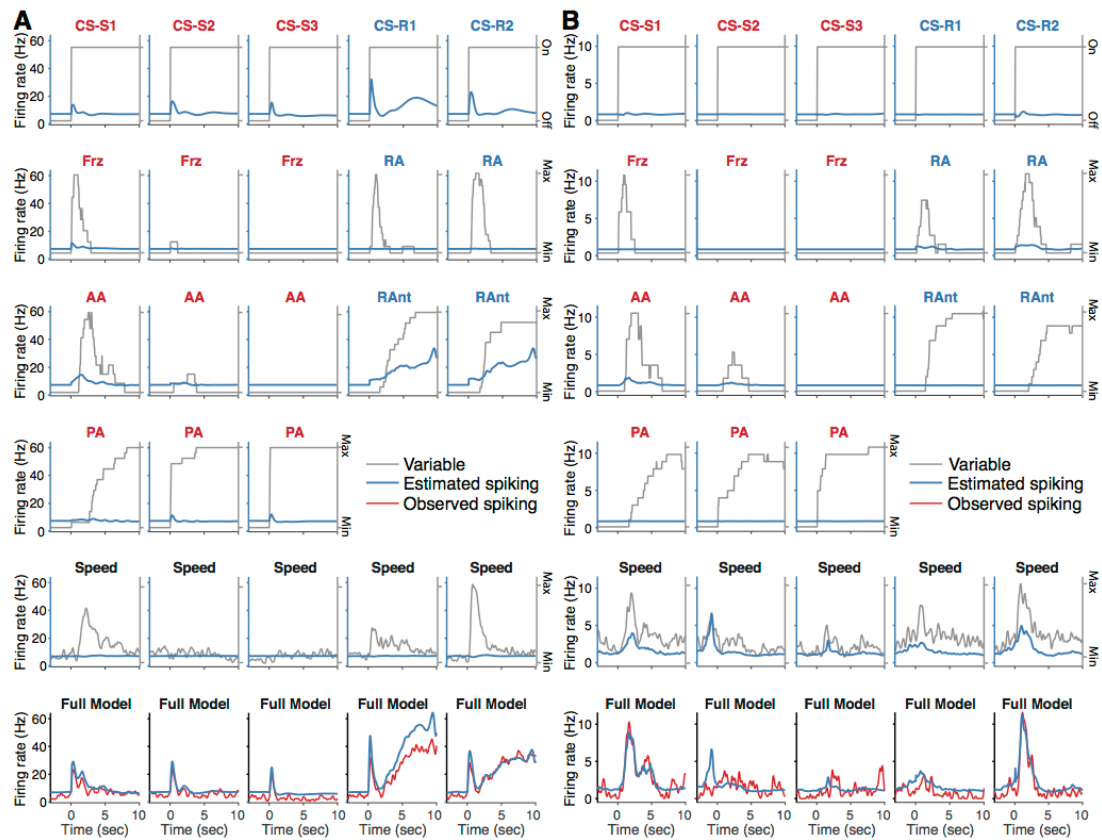


Figure 9. Coding of task variables by example BLA cells, as estimated by the GLM.

A and **B**, two different principal BLA cells. In both cases, the **six top rows** show GLM-estimated spiking (blue lines and left y-axis) for different task variables (gray lines and right y-axis) whereas the **bottom row** superimposes the observed spiking of the cell (red lines) and estimated spiking (full model, blue lines) for each CS. CS-Ss and associated behaviors (red letters) are shown on the left whereas CS-Rs and associated behaviors (blue letters) are shown on the right of each panel. Abbreviations: AA, active avoidance; Frz, freezing; PA, passive avoidance; RA, reward approach; RAnt, reward anticipation.

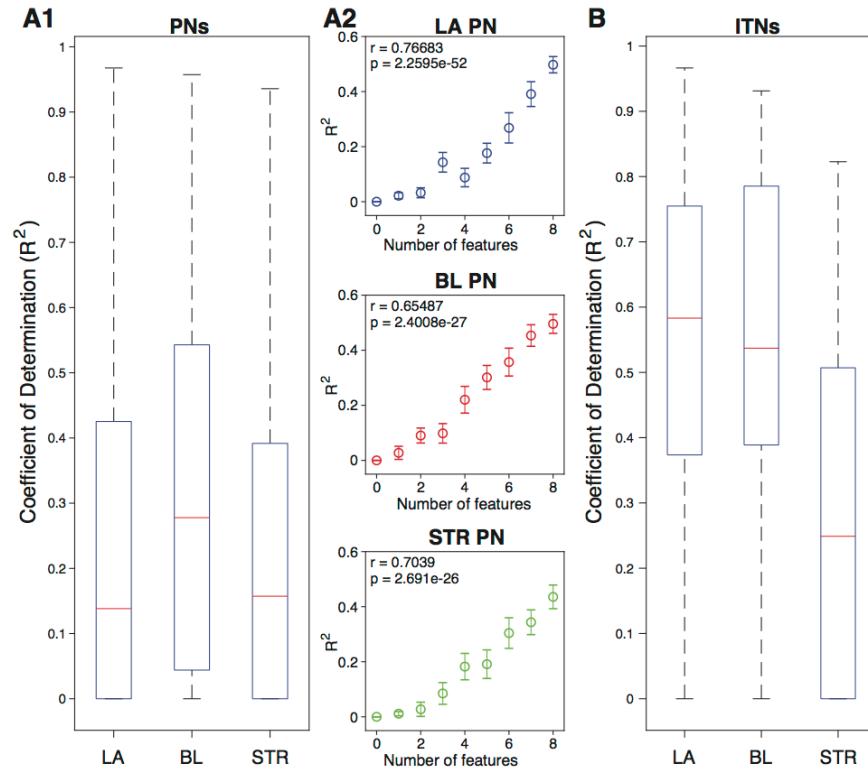


Figure 10. Coefficient of determination for principal cells and interneurons.

(A1) Boxplots of R^2 values for LA, BL, and striatal principal cells. Red line represents median value, lower and upper bounds of blue box represent 25th and 75th percentiles, respectively, and the whiskers extend to the most extreme data points not considered outliers. (A2) R^2 values as a function of the number of features encoded by each cell (Mean \pm SEM). Correlation coefficients and p-values reported on the top left corner of each plot. (B) Boxplots of R^2 values for interneurons.

Although the model estimates were generally consistent with the results of the standard analyses, there were some discrepancies. These likely resulted from the GLM's

ability to disentangle coding for concurrent variables that PETHs could not isolate. A prime example of this is the encoding of movement speed vs. active CRs (reward approach and active avoidance). This is illustrated in **figure 9B**, which shows a principal BLA cell that had been identified as an AA- and RA-cell with PETHs. In order for the model to assess whether the increases in firing rates seen during active CRs were, in sum or in part, actually due to the coding of locomotion, we had the GLM fit the cell's activity to the rat's movement speed throughout the task (including the inter-trial intervals). The model estimated that part of the spiking observed during active avoidance and reward approach trials was related to speed (**Fig. 9B** row 5). However, since speed did not fully account for the cell's increases in firing rates, the model attributed the rest of the response to the active CRs. Most importantly, the model-estimated spiking combined across all task-related variables closely matched observed spiking (**Fig. 9B**, bottom row). See **Fig. 10** for a characterization of the model's estimation accuracy.

3.6 Task-related activity in the BLA and striatum

Next, we used the GLM output to characterize coding of task variables by LA and BL neurons at the population level. For comparison, we carried out the same analyses in striatal neurons (n=208) recorded dorsal to the amygdala. In the three structures, $\geq 62\%$ of presumed principal cells (**Fig. 11A1**) and $\geq 65\%$ of interneurons (**Fig. 11A2**) encoded two or more task variables. The number of encoded task features was distributed similarly across structures in principal cells and interneurons (Kullbeck-Leibler tests for permuted distributions: all percentiles within 95th two-sided cut-offs; see General Methods).

To study regional differences in the coding of task variables, we computed the proportion of cells with excitatory and inhibitory modulations (respectively, red and blue bars in **Fig. 11B1,C1**) and their average absolute magnitude (**Fig. 11B2,C2**). These analyses revealed a remarkable degree of regional similarity in the encoding of task features. In the three regions, >48% of principal cells (**Fig. 11B1**) and >71% of interneurons (**Fig. 11C1**) showed excitatory or inhibitory modulations by all task features with the exception of freezing in the striatum (41% of PNs, 37% of ITNs). A three-way ANOVA on the magnitude of the modulations revealed that PNs (**Fig. 11B2**) had stronger modulations than interneurons (**Fig. 11C2**) across all task features ($F_{(1,3785)} = 71.52$, $p=0$). Furthermore, an interaction effect between feature type and structure

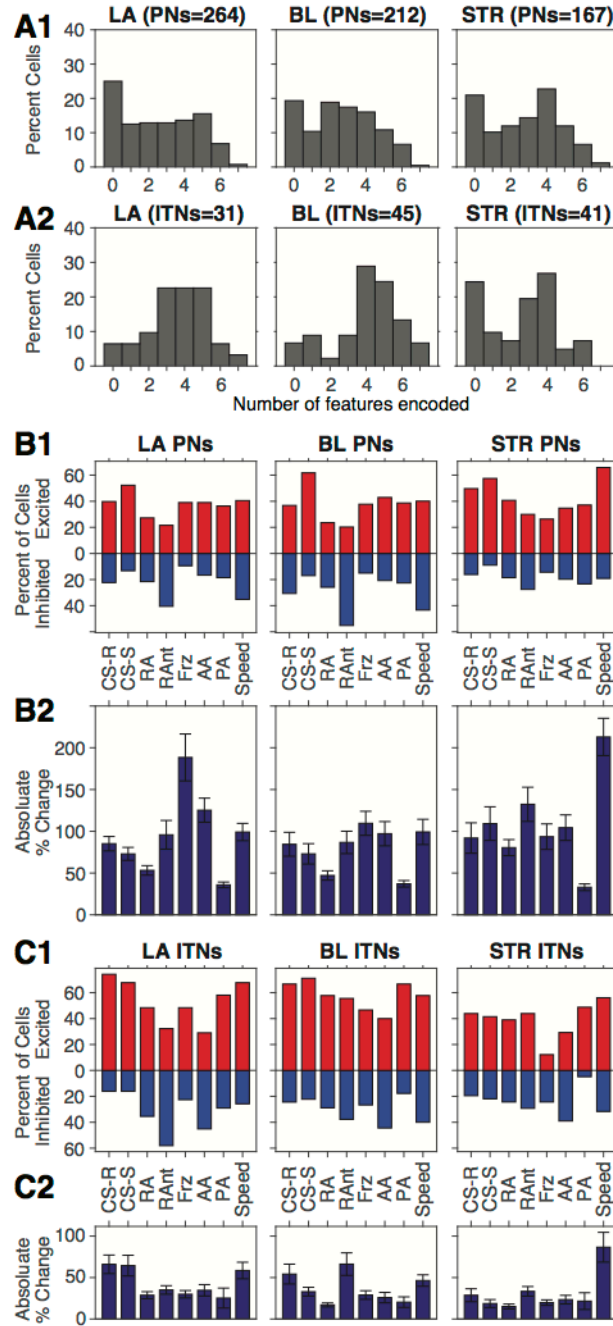


Figure 11. Multidimensional coding by BLA neurons, as determined by the GLM.

In A-C, the left, middle, and right columns show data obtained in LA, BL, and striatum, respectively. (A) Frequency distributions of the number of task variables encoded by presumed principal cells (A1) and interneurons (A2). (B1) Proportion of presumed principal cells (y-axis) that exhibited excitatory (red) or inhibitory (blue) coding of different task variables (x-axis). (B2) Absolute average \pm SEM modulation of firing rates in relation to each variable. (C1) Proportion of presumed interneurons (y-axis) that exhibited excitatory (red) or inhibitory (blue) coding of different task variables (x-axis). (C2) Absolute average \pm SEM modulation of firing rates in relation to each variable.

Abbreviations: AA, active avoidance; Frz, freezing; ITNs, interneurons; PA, passive avoidance; PNs, principal neurons; RA, reward approach; RAnt, reward anticipation.

($F_{(7,3785)}=2.47$, $p=0.0158$) indicated that striatal cells were more strongly modulated by speed than LA and BL cells (**Fig. 11B2**; Tukey-Kramer test, STR vs. LA $p=5.91 \times 10^{-6}$, STR vs. BL $p=5.91 \times 10^{-6}$), consistent with the role of the striatum in motor control (Brown and Robbins, 1989; Kravitz et al., 2010; Kreitzer, 2010; Reading et al., 1991). Finally, LA cells were most strongly driven by freezing and this modulation was significantly higher than in striatal but not BL neurons (**Fig. 11B2**; Tukey-Kramer test, LA vs. STR $p=0.0198$, LA vs. BL $p=0.0892$).

It was previously reported that BLA cells encode valence during appetitive and aversive tasks (Belova et al., 2008; Beyeler et al., 2016; Lee et al., 2017; Namburi et al., 2015; Sangha et al., 2013). In these prior studies however, the definition of valence was based on CS responsiveness, raising the question of the degree to which BLA cells encode valence in both the stimulus and behavioral dimensions. Presumably, if a cell is encoding valence, it should not only respond in a selective manner to the cues that elicit positive or aversive outcomes but also show corresponding changes in activity in relation to the conditioned behaviors they elicit. Thus, we defined valence as an excitatory response to either an appetitive or aversive CS, and at least one behavior of the same valence, plus an inhibition or no change in firing rate to the CS and behaviors of the opposite valence.

To study valence coding among BLA neurons, we plotted the *observed* normalized change in firing rate of LA and BL cells to reward approach (**Fig. 12**, y-axes) as a function of that seen in relation to active avoidance (**Fig. 12**, x-axes) and color-coded

the points corresponding to each cell based on the *model-estimated* modulation (blue, inhibition; red, excitation) by the various task variables (**Fig. 12**: top CSs; middle, active CRs; bottom, passive CRs). First, we verified that the model-estimated modulation by reward approach and active avoidance matched the change in firing rate observed in relation to these behaviors. Generally, the two matched: cells with strong positive estimated modulations by reward approach (red dots in **Fig. 12A2**) or active avoidance (red dots in **Fig. 12B2**) had higher values on the y- and x-axes, respectively (RA $r=0.56$, $p=1.75 \times 10^{-17}$; AA $r=0.66$, $p=3.29 \times 10^{-25}$). Cells that deviated from this trend strongly encoded speed, as discussed above (not shown).

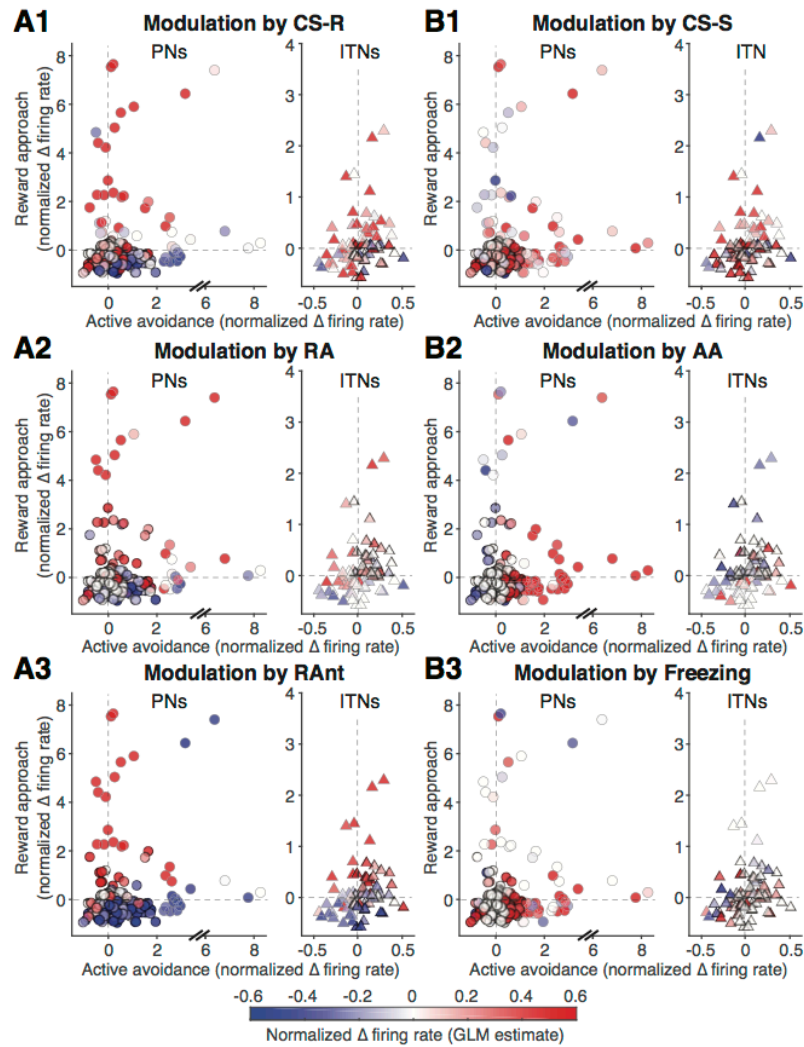


Figure 12. Relation between GLM-estimated coding of task variables and firing rate change observed during active conditioned behaviors.

For all available principal cells (left) cells or interneurons (right), all panels plot the peak firing rate change observed during reward approach (y-axis) as a function of that seen during active avoidance (x-axis). Although the position of the circles representing each cell does not vary between plots, their color changes, reflecting the degree to which they were, according to the GLM, excited (red) or inhibited (blue) in relation to the following variables: (A1) CS-R; (A2) reward approach - RA; (A3) reward anticipation -RAnt; (B1) CS-S; (B2) active avoidance - AA; (B3) freezing - Frz. See color scale at bottom of figure.

Consistent with the notion of valence coding in the BLA, a minority of principal BLA cells encoded valence along both stimulus and behavior dimensions. We first provide a qualitative description of these findings and then summarize the results of the statistical analyses. In **figure 12**, cells encoding positive valence are huddled near and high along the y-axis. They tend to show strong positive modulations by the CS-R (**Fig. 12A1**), reward approach (**Fig. 12A2**) and reward anticipation (**Fig. 12A3**). Conversely, cells encoding negative valence hug the x-axis. They tend to show strong positive modulations by the CS-S (**Fig. 12B1**), active avoidance (**Fig. 12B2**) and freezing (**Fig. 12B3**). Additionally, a subset of these valence-coding cells flipped their responses from an excitation to an inhibition in relation to variables of the opposite valence.

Overall, a low proportion of principal cells encoded positive or negative valence in LA (3.4 and 9.5%) and BL (1.9 and 17.5%), respectively. To test if this incidence was significantly higher than expected by chance given the proportions of cells whose activity was modulated by the various CSs and CRs, we permuted the normalized peak values of all task features (see Methods) 10,000 times to break any relationships between stimuli and behaviors and tested whether the proportion of valence-coding cells fell outside the two-sided 95th percentile of the null distribution. The incidence of valence-coding cells

did not exceed chance levels in LA or BL (33rd and 84th percentile, respectively). Of note, this negative result was not due to the fact that BLA neurons responded sparsely: most encoded multiple task features (**Fig. 11A**), including CSs and CRs (**Fig. 11B**). The low incidence of valence-coding we report is due to the fact that our definition of valence-coding not only takes into account the CS-specificity of the cells responses, as in prior studies, but also their behavioral correlates.

3.7 Heterogeneous Coding in LA and BL

Since explicit coding of valence is not prominent at the single cell level in the BLA, we tested if a valence code is present in the population activity. To examine this, for each region and cell type, we correlated the cells' peak modulations of all task variables (**Fig. 13**, dots mark significant values at $p < 0.001$). Principal LA neurons exhibited a marked tendency to show correlated responses to task events of the same valence, and decreased or negative correlations between those of different valences. This pattern was less striking for BL, but still apparent. To test the possibility that the correlations might partly result from correlations between the features themselves, we reassessed the significance of the correlation matrices by comparing them to shuffled neural activity. To this end, the peak values of each task features were shuffled 10,000 times and the correlation matrices were recomputed for each shuffled distribution. The actual correlation coefficients were considered significant if they fell outside the two-sided 99.9th ($p < 0.001$) percentile of the null distribution. However, the two approaches yielded nearly identical results: There was a 98% agreement between the two methods for principal cells and 100% for interneurons.

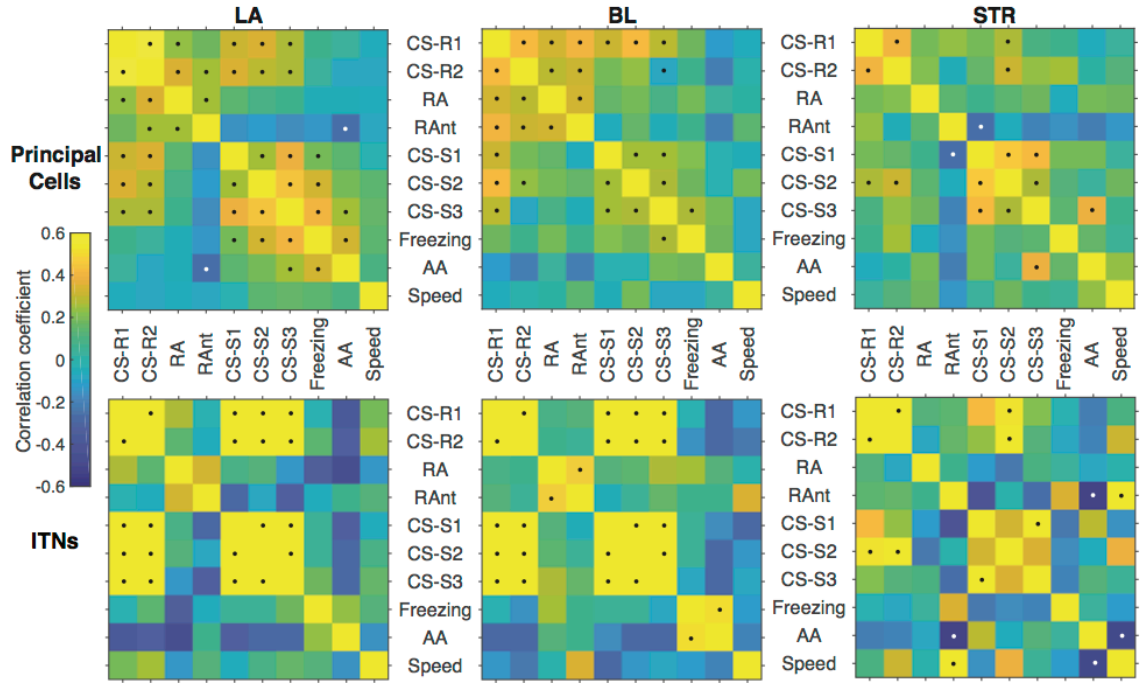


Figure 13. Correlation between firing rate modulations associated with behaviors and CSs.

Spearman correlation matrices for LA (left), BL (middle) and striatal (right) principal cells (top) or interneurons (bottom). Warmer and cooler colors indicate positive and negative correlations, respectively, as indicated by the color scale on the left. Dots indicate significant correlations ($p < 0.001$). Abbreviations: AA, active avoidance; Frz, freezing; PA, passive avoidance; RA, reward approach; RAnt, reward anticipation.

Similar results were obtained using a d-prime score (d' ; Keene et al., 2016; McKenzie et al., 2014) that measured the degree to which within-valence (CSs plus CRs) correlations exceeded those between-valence (bootstrap confidence intervals from zero, LA=1.56, $p < 10^{-4}$; BL=0.99, $p = 0.0009$). In LA, valence coding was also evident when restricting the d-prime analysis to the valence of behaviors ($d' = 5.99$, $p < 10^{-4}$) or CSs ($d' = 1.26$, $p = 0.004$), whereas in BL it was significant for behaviors ($d' = 1.96$, $p = 0.0009$) but not CSs ($d' = 0.58$, $p = 0.1$). The same analyses applied to ITNs revealed significant valence coding for behaviors (LA=3.27, $p = 0.014$; BL=6.57, $p = 0.0028$), but not CSs (LA=2.48, $p = 0.27$; BL=0.09, $p = 0.4$).

While the above analyses indicate that valence is coded along multiple dimensions in the BLA, it remains unclear how these codes are distributed at the population level and how they relate to each other. To address these questions, we applied multidimensional scaling (MDS; **Fig. 14A**), a dimensionality reduction strategy that proved useful to study coding in the gustatory and olfactory systems, where sensory representations are obscure (Di Lorenzo, Chen, & Victor, 2009; Youngentob, Johnson, Leon, Sheehe, & Kent, 2006). The peak firing rate modulations related to the ten variables of interest (3 CS-Ss, 2 CS-Rs, 4 CRs, and speed) were used to plot the position of cells in a ten-dimensional space (**Fig. 14A1**). This map was then collapsed into two dimensions (**Fig. 14A2,3**) based on the similarity of the cells' peak responses, while minimizing errors in setting their relative distance. Once this map was generated for each region, we examined how it related to various coding dimensions (e.g. valence-behavior, valence-CS) by fitting a plane in 3D space where the first two dimensions were provided by the MDS and the third was one of the coding dimensions (**Fig. 14A4,5**; labels at top of plots in **Fig. 14B**). In these maps, the steepness of the color gradient is proportional to how strongly the coding dimension is represented in the population. The orientation of the color gradients indicates whether the different coding dimensions are related. That is, orthogonal orientations indicate independent codes whereas parallel orientations indicate closely related codes. For simplicity, **figure 14C** represents the steepness and orientation of the color gradients by vectors, one for each of the coding dimensions examined. For consistency, in each region, we set to vertical the vector associated with the coding dimension most strongly represented, while preserving the relative orientation of the other vectors.

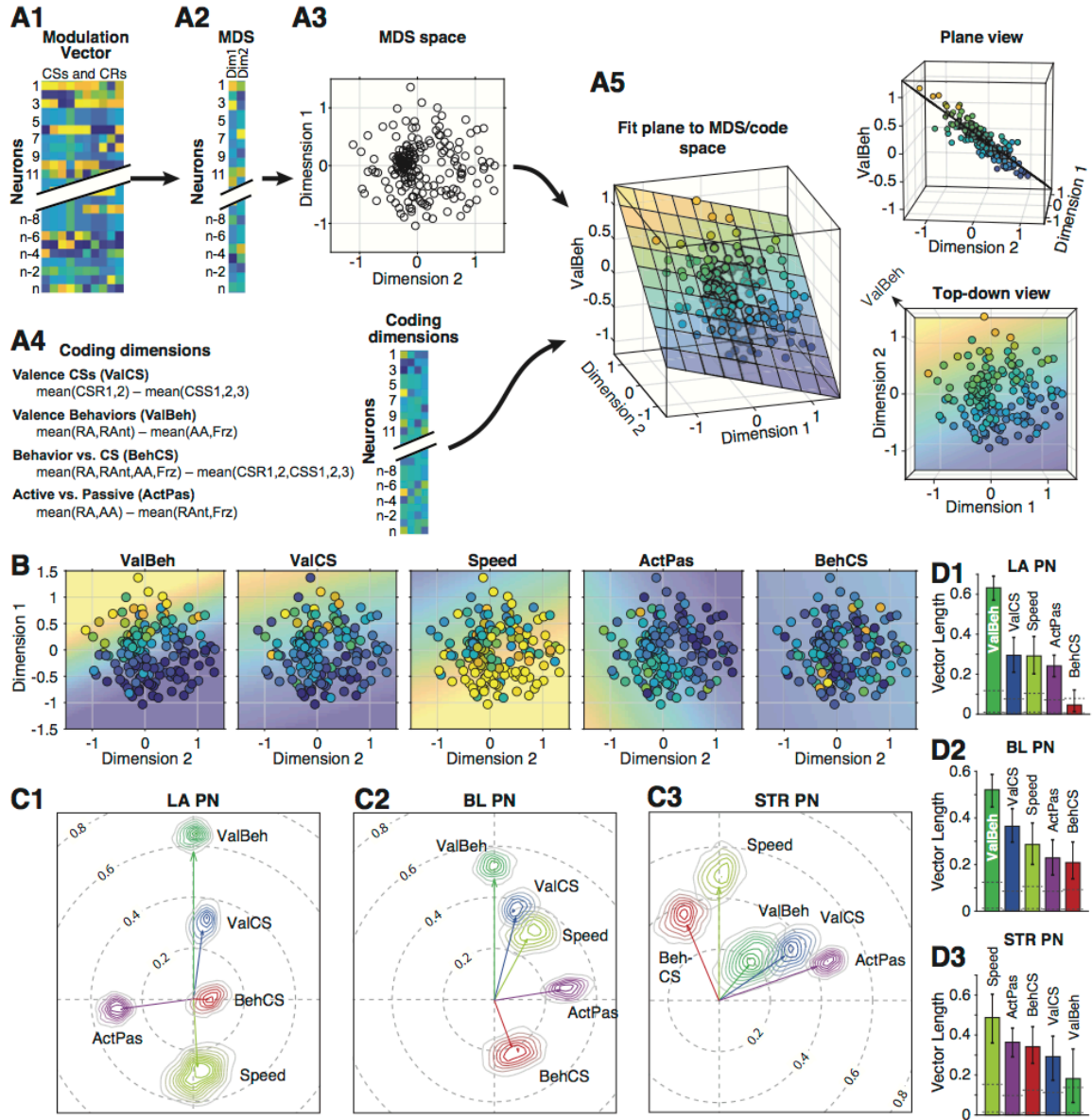


Figure 14. Coding for different task dimensions at the population level.

(A1) Each neuron was described by a vector composed of their responses to CSs and behaviors. (A2,3) These were mapped to a two dimensional space using MDS. (A4) For each neuron we also computed a code value that was derived by contrasting responses to different stimuli and behaviors. (A5) PNs from each region were placed in a space where the first two dimensions were their MDS values, and the third was the value for one of their codes. A three dimensional plane was fit in this space that could capture a systematic mapping of the code under consideration in the low dimensional space. (B) Low dimensional maps of coding for LA PNs. Each PN (filled circle) was placed into a two dimensional space such that nearby PNs had similar response vectors. The code value for each neuron was then added (color of the filled circle). A code that is systematically represented in the low dimensional response space produces a strong gradient (e.g. ValBeh), and one that is not produces a weak gradient (e.g. BehCS). (C)

Length and direction of plane gradients from the PNs recorded in each region. The gradient for each plane plotted on a polar plot as colored arrows. The plot was rotated so that the gradient with the longest length pointed up at 90 degrees. Contour plots with fading colors are the bootstrap probability distributions for the gradients. **(D)** Ordering of the coding gradients by length. Error bars denote the 95% bootstrap confidence intervals for the gradient length. Dashed lines are 95% confidence intervals returned by computing null gradients where code value was randomly permuted across the population.

Consistent with the correlation matrices (**Fig. 13**), valence-behavior was the strongest coding dimension in LA and BL (**Fig. 14C1,2**). Although less salient, the valence-CS dimension was also expressed in LA and BL and aligned to valence-behavior. Coding for speed also fell roughly along the same axis as valence, except that its direction was opposite in LA and BL. Evidence of this can also be found in the correlation matrices where the speed-related activity of LA neurons tended to correlate positively with their modulation by aversive events (**Fig. 13**, top left) and less so or inversely in BL (**Fig. 13**, top middle). The active-passive coding dimension was also present in LA and BL (**Fig. 14C1,2**), but it was orthogonal to the valence code, implying that the two populations could simultaneously represent both types of information with minimal interference. As for speed, the modulation along the active-passive coding dimension was opposite in LA vs. BL. Finally, while the same coding dimensions were observed in the striatum (**Fig. 14C3**), their relative importance and overall organization differed markedly from LA and BL. In particular, speed emerged as the dominant coding dimension followed by the active-passive, behavior-CS, valence-CS, and valence-behavior codes.

To assess the variability of vector length estimates, we used a 95% bootstrap confidence interval (**Fig. 14D**). For each vector length, we also defined 95% confidence intervals returned by computing null gradients where code values were randomly

permuted across the population (dashed lines in **Fig. 14D** denote 97.5 and 2.5 percentile of the null distributions). Using this approach, it was determined that all vector lengths were significant with the exception of behavior-CS in LA (**Fig. 14D1**) and valence-behavior in the striatum (**Fig. 14D3**). To assess whether vector lengths differed significantly from the others within a region, we used a permutation test with correction for multiple comparisons, confirming that valence-behavior was the dominant coding dimension in LA and BL ($p=0.04$; **Fig. 14D1,2**) whereas speed dwarfed the others in the striatum ($p=0.04$; **Fig. 14D3**). Bootstrap resampling with threshold p of 0.05 (corrected for ten comparisons) was also used to assess the difference of angle estimates between codes (**Fig. 14C**, concentric circles at tip of vectors), indicating that the active-passive coding dimensions differed significantly from the valence dimensions in LA and BL (valence-behavior vs. active-passive BL $p=0.04$, LA $p<0.002$; valence-CS vs. active-passive BL $p=0.04$; LA $p<0.002$), but not in the striatum (both $p>0.05$). See **Fig. 15** for similar analyses in interneurons.

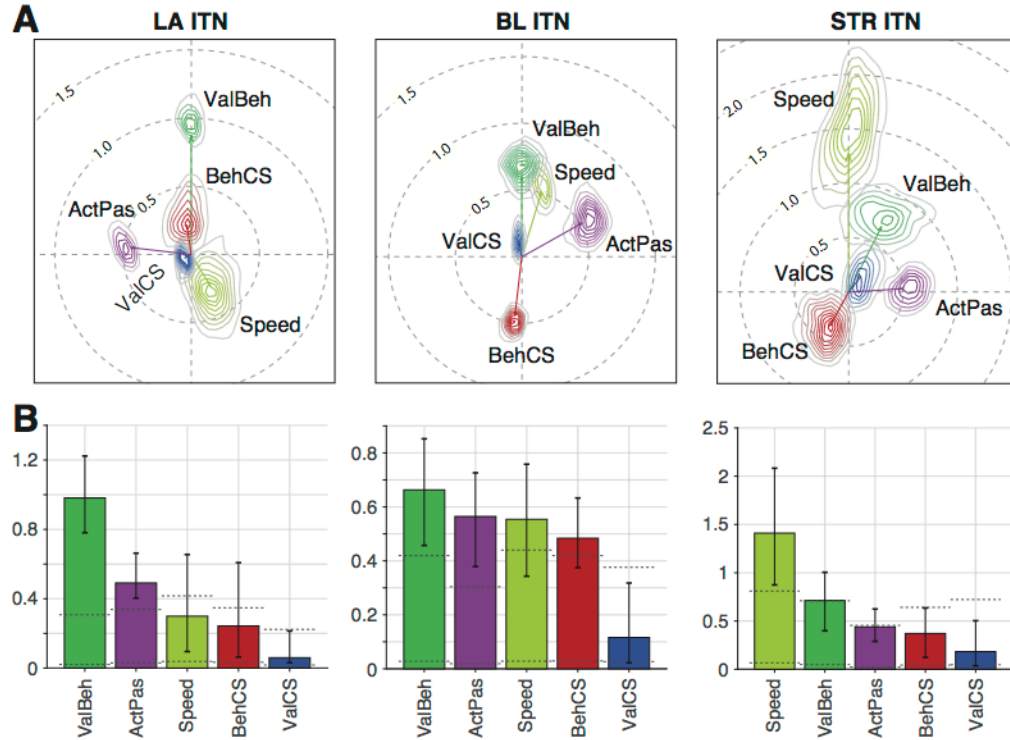


Figure 15. Putative interneurons in both the amygdala and striatum exhibited systematic mapping of coding.

(A) Coding gradients for LA, BL, and STR interneurons plotted similarly to those seen in Fig. 14C. In both LA and BL, ValBeh had the strongest gradient, although it was only significantly stronger in LA when compared with ValCS ($p = 0.04$). Interneurons in STR showed the strongest coding for speed, which was significantly stronger than the codes for ActPas ($p = 0.04$), ValCS ($p = 0.04$), and BehCS ($p = 0.04$). (B) Coding strengths for interneurons across LA, BL, and STR plotted similarly to those seen in Fig 14D.

3.8 Conclusions

It was reported that appetitive or aversive conditioning leads to the potentiation of CS inputs onto valence-specific BLA neurons (McKernan & Shinnick-Gallagher, 1997; Rumpel et al., 2005; Tye et al., 2008). As a result, CSs would trigger approach or defensive CRs through the activation of valence-coding neurons with distinct outputs (Beyeler et al., 2016; Namburi et al., 2015; Tye et al., 2008). However, because of the close temporal relation between CSs and CRs, and the fact that in most prior studies, each CS could only trigger one CR, it remained unclear whether BLA neurons encode CS

identity, the behavior it elicits, or if the two are one and the same. Our study aimed to shed light on this question using a task where rats could emit different CRs in response to the same CSs, allowing us to dissociate normally intertwined task features that BLA neurons potentially encode.

Using this approach, we found that in most LA and BL cells, the magnitude and time course of their activity during the CS depended on the type and timing of the CRs. Thus, the CS responses of BLA cells, that is their sensory aspects, are separable from the activity that drives CRs; one does not necessarily cause the other. Second, while valence could be decoded at the population level, the incidence of valence-coding cells, as defined by their corresponding CS- and CR-related selectivity, did not exceed chance. Last, most BLA neurons concurrently encode multiple task features and behaviors.

Chapter IV:

Multidimensional representations in the amygdalo-prefrontal network

4.1 Rationale

Of the structures regulating memory and emotions, two nodes stand out for their remarkably similar connectivity: the medial prefrontal cortex (mPFC) and amygdala (reviewed in Ongur & Price, 2000; Pitkanen, 2000). Not only do these structures form dense reciprocal connections with each other, but they both have access to associative sensory information, share close ties with the insula, project to overlapping striatal territories, and target a common array of brainstem nuclei involved in neuromodulation, emotional expression, and autonomic regulation (Allen et al., 1991; Gabbott et al., 2005; Hoover & Vertes, 2007; Krettek & Price, 1977, 1978; McDonald, 1998; Ongur & Price, 2000).

Consistent with their similar connectivity, the functions of the amygdala and mPFC are tightly intertwined. For instance, they are both required for the expression and extinction of conditioned responses (CRs) to cues (CSs) that predict aversive outcomes (USs) such as behavioral freezing and active avoidance (Bravo-Rivera et al., 2014; Fanselow & Poulos, 2005; LeDoux, 2000; Moscarello & LeDoux, 2013; D. Sierra-Mercado, Jr. et al., 2006). Moreover, they both regulate cued reward-seeking behaviors

(Ambroggi et al., 2008; Burgos-Robles et al., 2013; Ishikawa et al., 2008; Peters et al., 2009).

If the roles of the amygdala and mPFC overlap so much, what is the computational advantage of having both? One possibility is that mPFC and amygdala neurons code information differently, allowing them to implement distinct sets of input-output functions. Thus, in this chapter I compare the coding properties of mPFC and amygdala neurons using the RRI task.

4.2 CS responsiveness and behavioral correlates of PL neurons

Rats were trained on the RRI task, as described in the previous chapter, to avoid footshocks based on the sector where the CS-S appeared, and retrieve rewards based on the location of the CS-R (**Fig. 16A**). Once they reached criterion (>85% correct trials on all CSs combined), generally in less than a week (**Fig. 16B**), they were implanted with a silicon probe just dorsal to the prelimbic (PL) sector of the mPFC (n=4). Three of these rats were also implanted with silicon probes in the amygdala. Throughout the recording sessions, we monitored the rats' position, head direction, and movement velocity with an overhead camera. The probes were lowered after each recording session by at least 140 μm to obtain new cells the following day. At the conclusion of the experiments, the location of the probe was marked with small electrolytic lesions for subsequent histological verification of recording sites (**Fig. 16C**).

Cells recorded in the amygdala in chapter 3, were combined for analyses with all newly recorded neurons. Only cells histologically determined to have been recorded in PL (n=526), LA (n=365), and BL (n=307) are included in this chapter. PL and amygdala

cells were classified as putative PNs or fast-spiking interneurons based on their firing rates and spike duration (trough-to-peak interval; **Fig. 16D**). However, only PNs will be considered further, resulting in samples of 452 cells in PL, 284 cells in LA and 214 BL neurons.

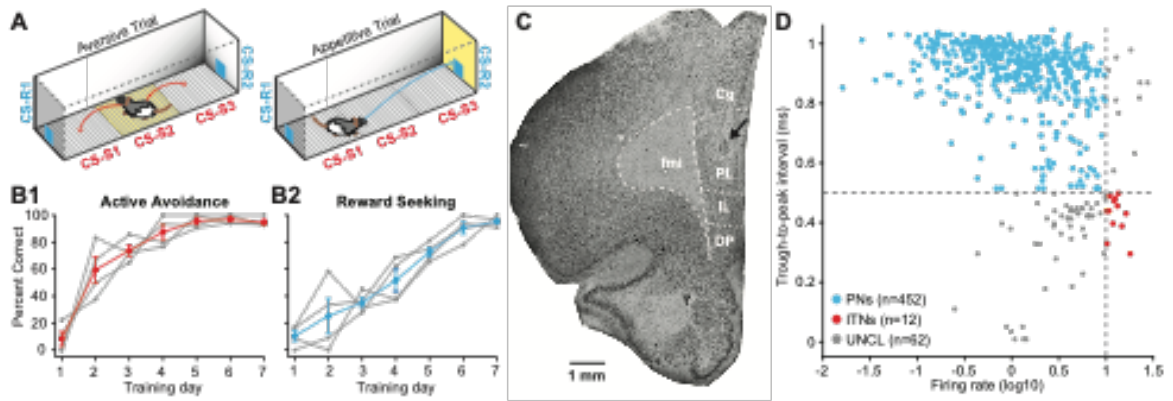


Figure 16. RRI task, histological verification of PL neurons, and their classification

(A) RRI task apparatus: a rectangular arena with high walls, no ceiling, and a floor made of metal bars. Light-emitting diodes (LEDs) at different positions signal an upcoming reward (blue; behind left and right wall, CS-R1 and CS-R2, respectively) or an impending footshock (red; under each of three different floor sectors, CS-S1-3). (B) Acquisition of conditioned active avoidance (B1) and reward seeking (B2). Gray lines, individual rats ($n=4$). Colored lines, average \pm SEM. (C) Histological verification of recording site. Coronal section stained with cresyl violet. A small electrolytic lesion (arrow) marks the last recording site. (D) Classification of recorded cells as presumed projection cells (PNs, blue) or fast-spiking interneurons (ITNs, red) based on their spike duration (y-axis, peak-to-trough interval, cut-off of 0.5 ms) and firing rate (x-axis, cut-off of 10 Hz). Cells that did not meet both criteria (UNCL, gray) were not considered further. Abbreviations: Cg, cingulate cortex; DP, dorsal peduncular cortex; fmi, forceps minor of the corpus callosum; IL, infralimbic cortex; PL, prelimbic cortex.

The task-related activity of PL neurons was similar to that of LA and BL cells in many respects. **Figure 17** illustrates the responses of LA, BL, and PL neurons to CS-Rs (top) and CS-Ss (bottom), z-scored based on their activity during the pre-CS period (5 s). Cells were rank-ordered based on the amplitude of their responses to CS-R1 or CS-S1 and the same order was kept for CS-R2 and CS-S3, respectively. Although the magnitude

of firing rate changes tended to be lower in PL (see also below), CS responses at the three sites were similar, ranging from transient or sustained increases in firing rates to long lasting inhibitory responses (**Fig. 17**).

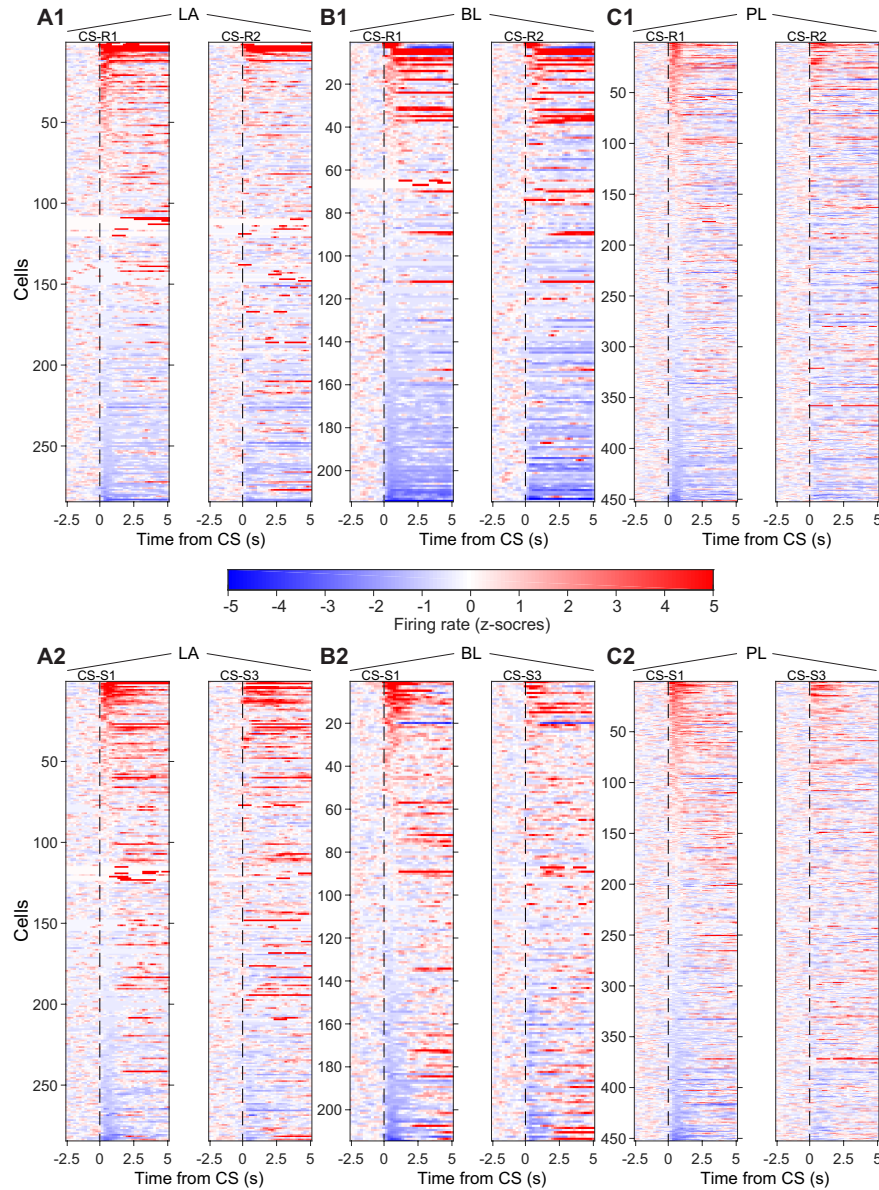


Figure 17. CS-evoked activity in LA, BL, and PL neurons.

Heatmaps of activity evoked by CS-Rs (top) or CS-Ss (bottom) in all available principal LA (**A**, $n=284$), BL (**B**, $n=214$), and PL (**C**, $n=452$) neurons. Firing rates were z-scored based on activity during the pre-CS period. Data is plotted with 200 ms bins. Warmer colors indicate higher firing rates (see color bar). In **A1-C1**, neurons were rank ordered based on the amplitude of their responses to CS-R1 and the same order was kept for CS-

R2, respectively. In **A2-C2**, cells were rank ordered based on the amplitude of their responses to CS-S1 and the same order was kept for CS-S3.

To assess the statistical significance of CS responses, we compared firing rates during the time bins in the first second following CS-onset to the 5-s pre-CS baseline (50 ms bins) using rank-sum tests with a significance threshold of $p < 0.005$. We restricted this comparison to the first second after CS onset because other analyses described below revealed that PL neurons, like BLA cells, also fired in relation to CRs occurring later during the CSs. With this approach, the proportion of cells with significantly increased firing rates during at least one of the CSs did not differ significantly at the three sites for the CS-Rs (LA, 7%; BL, 9%, PL, 7%; $\chi^2 = 0.31$, $p = 0.86$) or the CS-Ss (LA, 8%; BL, 13%, PL, 11%; $\chi^2 = 3.36$, $p = 0.19$; **Fig. 18**).

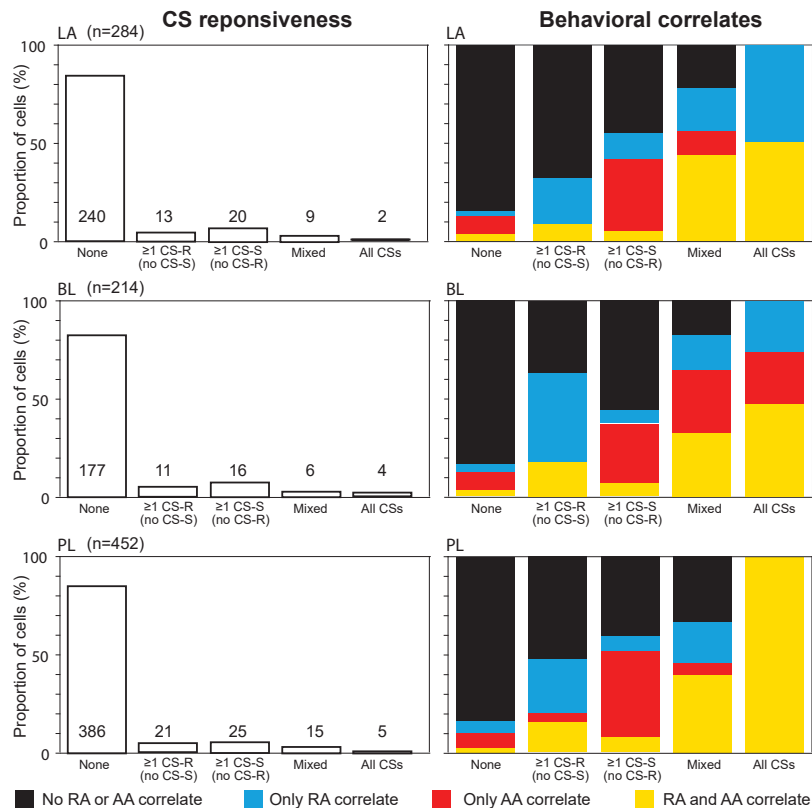


Figure 18. CS and behavior response patterns of LA, BL, and PL neurons.

Percent of cells with a significant increase in firing rate (rank-sum test, $p=0.005$) during time bins in the first one second after CS or behavior onset compared to 5 s baseline pre-CS period in LA, BL, and PL. Left: Percent of cells with no CS response, response to at least one CS-R and no CS-S response, response to at least one CS-S and no CS-R response, mixed responses to either CS-R or CS-S, and all CSs. Right: Percent of cells with one of the CS-response categories on the left and an increase in firing rate during AA, RA, neither or both.

Another similarity between LA, BL and PL neurons was the strong dependence of their CS-related activity on behavior. To examine this question, we referenced unit activity to the onset of CRs instead of CSs, computed peri-event time histograms (PETHs) of neuronal discharges, and determined whether firing rates increased in relation to one or more CRs relative to firing rates preceding behavior onset (rank-sum tests, $p<0.005$; see Methods). **Figure 19** illustrates three different PL neurons where trials were aligned to the onset of CSs (top) or CRs (bottom). As we reported in LA and BL (Kyriazi et al., 2018), some PL cells with robust CS responses lacked behavior-related activity (**Fig. 19A**). Conversely, some PL cells with no CS responses showed strong CR correlates (**Fig. 19B**). And finally, some cells had both, CS- and behavior-related increases in firing rates (**Fig. 19C**). A detailed analysis of the relation between CS responsiveness and CR-related activity revealed a remarkably similar profile at the three sites (**Fig. 18**). The proportion of cells with increased firing rates during active CRs did not differ significantly between amygdala and PL neurons for the CS-R (RA: LA, 9.9%; BL, 12.6%; PL, 13.2%; $\chi^2=1.98$, $p=0.37$) or CS-S (AA: LA, 16.5%; BL, 17.2%; PL, 15%; $\chi^2=0.64$, $p=0.73$).

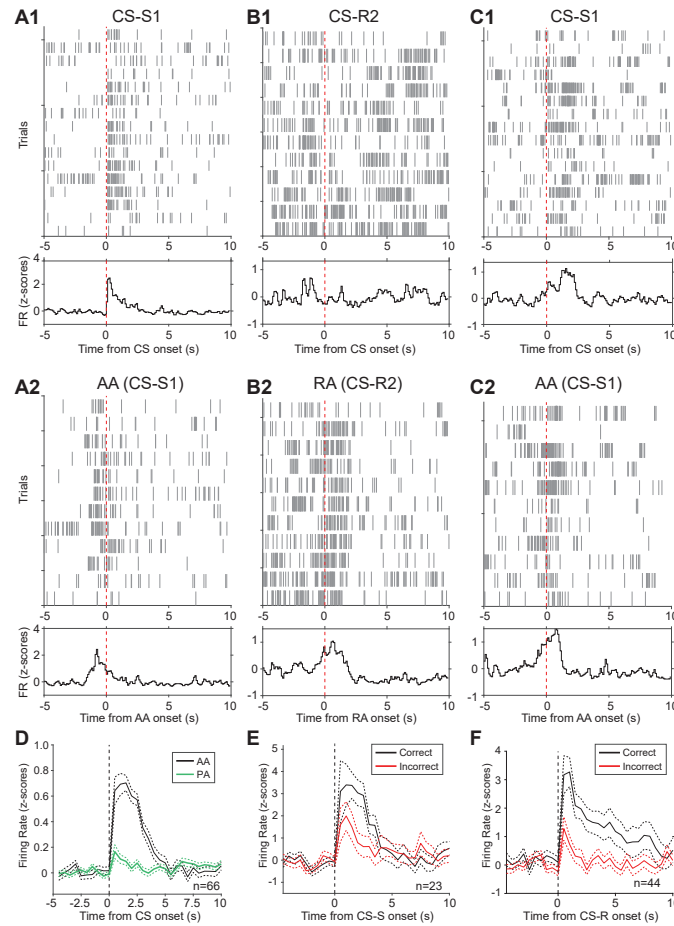


Figure 19. CS- and CR-related activity of PL neurons.

(A-C) Three different PL neurons. Top, rasters where ticks represent individual spike times and each row represents a trial. Bottom, z-scored peri-event histograms (PEHs) of firing rates. In A1-C1, rasters and PEHs are referenced to CS onset whereas in A2-C2, they are referenced to CR onset. (A) Cell with a CS response but no CR correlate. (B) Cell with no CS response that fires in relation to the CR. (C) Cell with both, a CS response and a CR correlate. (D) Comparison between z-scored averaged firing rate \pm SEM of PL cells during AA (red) vs. PA (blue), referenced to CS onset. Only cells increasing their firing rate during AA are included in this panel. (E) Comparison between z-scored average firing rate \pm SEM of PL cells during correct (black) and error (red) CS-S trials. The number of cells included in this analysis is lower than in panel E because error trials did not occur in all recording sessions. This figure is based on ten error trials from seven recording sessions. (F) Comparison between z-scored average firing rate \pm SEM of PL cells during correct (black) and error (red) CS-R trials (60 error trials from 12 sessions).

In fact, further analyses revealed that, as we previously observed in LA and BL (Kyriazi et al., 2018), much of the CS-related increases in PL firing rates, even those developing shortly after CS onset, were not linked to the CSs but to the behavior they elicited. For instance, when we compared the same cells with AA-related activity on CS-S trials that called for AA vs. PA responses, we found that the CS-S related activity was drastically lower on trials that called for PA than AA responses (**Fig. 19D**). Similarly, when we compared the activity of AA or RA cells on correct vs. error trials, their firing rate during the CS was lower on error than correct trials (**Fig. 19E-F**).

4.3 Differences between the selectivity of PL and BLA neurons

Despite the similarities, there were also clear differences. First, the cells' responses to different CSs of the same type (e.g. CS-R1 vs. CS-R2; CS-S1 vs. CS-S3) were more dissimilar in PL (**Fig. 17C**) than in LA (**Fig. 17A**) or BL neurons (**Fig. 17B**). To quantify this, we first averaged the z-scored response of each cell during the first second after CS onset and then calculated the correlation between CS-R1 vs. CS-R2 and CS-S1 vs. CS-S3 for each brain region, separately. The correlation was lower in PL than LA or BL neurons for the CS-Rs (LA, 0.53; BL, 0.69; PL, 0.42) and CS-Ss (LA, 0.5, BL, 0.58; PL, 0.45).

To shed light on the origin of this difference, we then compared the incidence of cells whose firing rates significantly increased or decreased in response to just one of the CSs of a given type. This method disclosed a clear difference between PL and amygdala neurons. That is, the proportion of cells with inhibitory responses to only one CS of a given type was significantly higher in PL (**Fig. 20A1**, blue). This was true of the CS-Rs

($\chi^2=18.7$, $p<0.0001$) and the CS-Ss ($\chi^2=9.64$, $p=0.0081$) but did not hold for excitatory responses (**Fig. 20A1**, red) to CS-Rs ($\chi^2=1.03$, $p=0.59$) or CS-Ss ($\chi^2=0.17$, $p=0.92$).

Figure 20A2 shows an example of cell with inhibitory response to CS-R2 but not to CS-R1.

PL neurons also showed greater selectivity for CRs evoked by different CSs. That is, a higher proportion of PL cells were activated in relation to RA evoked by only one of the CS-Rs (e.g. RA to CS-R1 but not CS-R2) or AA elicited by only one of the CS-Ss (e.g. AA to CS-S1 but not 3) than in LA or BL (RA: $\chi^2=12.11$, $p=0.0024$; AA: $\chi^2=19.99$, $p<0.0001$; **Fig. 20B1**, red). The difference in selectivity was less marked (RA) or absent (AA) when considering inhibitory responses (RA: $\chi^2=20.61$, $p<0.0001$; AA: $\chi^2=0.03$, $p=0.99$; **Fig. 20B1**, blue). **Figure 20B2** shows an example of a cell activated in relation to AA elicited by CS-S1 but not CS-S3.

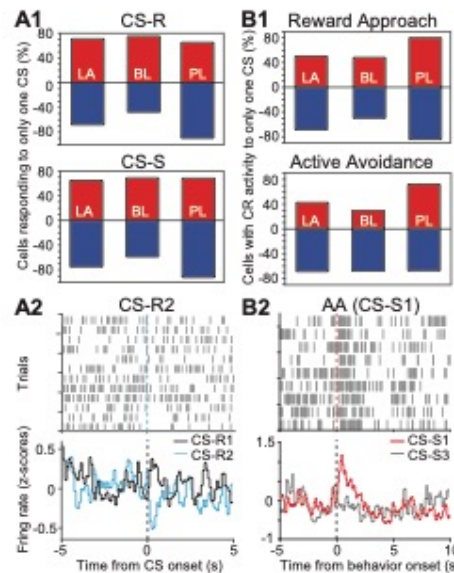


Figure 20. Higher incidence of CS- and CR-selective cells in PL

(A1) Comparison between the incidence of cells with excitatory (red) or inhibitory (blue) responses to only one CS-R (top) or CS-S (bottom) among LA (left), BL (middle) and PL (right) neurons. Percentages represent selective cells out of total number of responsive cells. The total number of CS-R excited cells was 24 in LA, 22 in BL and 40 in PL. The

total number of CS-S excited cells was 31 in LA, 26 in BL, and 47 in PL. The total of CS-R inhibited cells was 25 in LA, 52 in BL, and 49 in PL. The total number of CS-S inhibited cells was 20 in LA, 46 in BL and 38 in PL. **(A2)** Example of cell with inhibitory response to CS-R2 but not to CS-R1. **(B1)** Comparison between the incidence of cells with excitatory (red) or inhibitory (blue) CR-related activity in relation to only one CS-R (top) or one CS-S (bottom). The total number of RA-excited cells was 28 in LA, 27 in BL, and 60 in PL. The total number of AA-excited cells was 47 in LA, 37 in BL, and 68 in PL. The total number of RA-inhibited cells was 35 in LA, 44 in BL, and 148 in PL. The total number of AA-inhibited cells was 47 in LA, 37 in BL, and 68 in PL. **(B2)** Example of cell activated in relation to AA elicited by CS-S1 but not CS-S3. **(A2, B2)** Top, rasters where ticks represent individual spike times and each row represents a trial. Bottom, z-scored PEHs of firing rates.

4.4 Disentangling the correlates of unit activity with a generalized linear model (GLM)

In the prior section, our analyses were complicated by the presence of multiple sensory and behavioral variables whose relative timing varied between trials. Specifically, a number of factors potentially influenced neuronal activity, including CS identity, the rats' position with respect to the CS, the type and timing of the required (or erroneous) CRs, the speed at which the rats moved, and potential interactions between these variables. Since PSTHs cannot possibly isolate the influence of these interacting factors, we turned to a different approach: we inferred the variables PL neurons encode using a group lasso GLM with ten-fold cross validation.

The GLM relies on variations in the timing, duration and type of variables occurring on different trials to determine which ones neurons encode. Of note, the group lasso GLM permits dimensionality reduction in correlated data and favors sparsity in the identification of variables related to neuronal activity (Breheny & Huang, 2015; R. Tibshirani, 1996; M. Yuan & Y. Lin, 2006). Due to the prevalence of selective responses observed in PL (**Fig. 20**), we added interaction terms in the GLM to capture the differential associations between individual trial types and behaviors (e.g. CS-R1 with

RA, CS-R2 with RA, CS-S1 with AA, CS-S3 with AA, etc.; see Methods for full list of parameters).

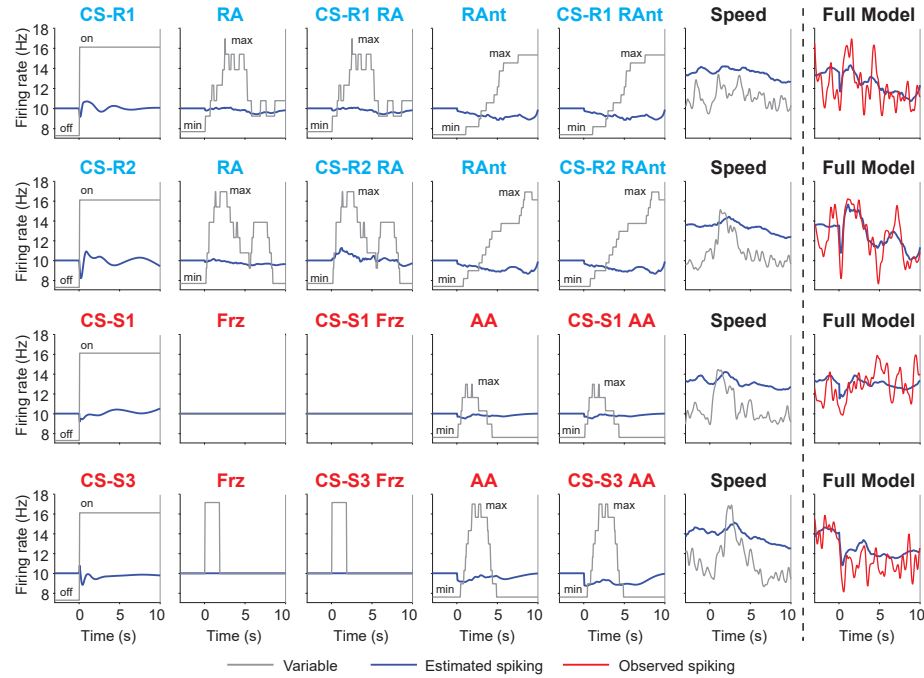


Figure 21. Coding of task variables by an example PL cell as estimated by the GLM. The first six columns show GLM-estimated spiking (blue lines) for each task variable (gray lines). The last column on the right consists of the estimated spiking (blue lines) superimposed on the observed spiking of the cell (red lines) for each CS. CS-Rs and associated behaviors (blue letters) are shown in the first two rows, CS-Ss and associated behaviors (red letters) shown in the bottom two rows. Abbreviations: AA, active avoidance; Frz, freezing; RA, reward approach; RAnt, reward anticipation.

To ensure the GLM captured the observed coding parameters, we compared the model's output to PETHs referenced to different variables. One example cell shown in **figure 21** had a complex profile of responses, including a transient inhibition at the onset of all CSs, a selective increase to RA only on CS-R2 trials, and an inhibition related to AA. As indicated in the first six columns of figure 5, the model accurately captured all of these responses, including selective responses with interaction terms such as CS-R2 RA (row 2, column 3). Additionally, it identified that a proportion of the response was related

to speed, a parameter that is tracked throughout the sessions. Locomotion effects on firing rate are a factor that traditional PETH analyses cannot account for. Often, apparent behavior correlates are in fact related to changes in the animal's speed (see example of such a cell in **figure 22**). The GLM allows us to factor speed out, and attribute firing related to behaviors without the confounding influence of movement speed.

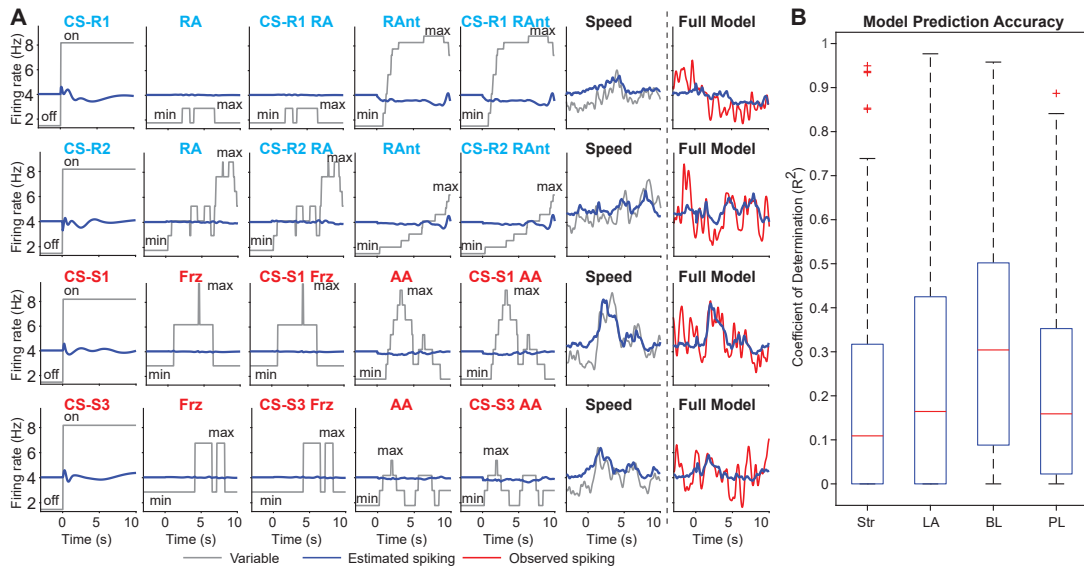


Figure 22. Example PL neuron with speed correlate fit with GLM and model accuracy.

(A) A PL cell with a speed correlate estimated by the GLM. The first six columns show GLM-estimated spiking (blue lines) for each task variable (gray lines). The last column on the right consists of the estimated spiking (blue lines) superimposed on the observed spiking of the cell (red lines) for each CS. CS-Rs and associated behaviors (blue letters) are shown in the first two rows, CS-Ss and associated behaviors (red letters) shown in the bottom two rows. Abbreviations: AA, active avoidance; Frz, freezing; RA, reward approach; RAnt, reward anticipation. (B) Coefficient of determination (R^2) for all putative principal cells recorded in striatum, LA, BL, and PL. Red lines represent median values, blue box bottom and top lines represent the 25th and 75th percentiles, respectively, black whisker lines extend to the most extreme data points. Red crosses represent outliers.

Another benefit of the lasso GLM is its ability to set parameters to zero when they do not contribute to the firing rate prediction of a given cell. For example, in **figure 21**,

the freezing behavior variables are all set to zero including both main and interaction effects, because the firing rate of the cell was unaffected by freezing even though the animal spent several trials freezing to CS-S1. This ability prevents overfitting and favors sparsity in identifying variables related to firing rate.

The GLM output was used to characterize coding by LA, BL, and PL neurons. As a control, we performed the same analyses on putative principal striatal cells ($n=237$) that were recorded dorsal to the amygdala (see identification criteria in Methods). The model's estimation accuracy was comparable between striatum, amygdala, and PL cells, although the fit for BL cells was superior to the other regions (**Fig. 22B**; $\chi^2(3)=25.81$, $p<0.0001$). Based on the model predictions, we computed the proportion of cells with inhibitory and excitatory modulations (**Fig. 23A**, blue and red bars, respectively) as well as the average absolute magnitude of the modulations (**Fig. 23B**). With few exceptions, the encoding of task variables was similar in the four regions. At the four sites and for each task feature, $>35\%$ of cells (**Fig. 23A**) showed excitatory or inhibitory modulations with the exception of freezing, a variable that was especially under-represented among PL neurons.

A two-way ANOVA with recording site and feature type as factors on the modulation of responses revealed a significant main effect for recording site and feature ($F_{(3,17)}=96.77$ and 41.8 , respectively; $p<0.0001$; **Fig. 23B**). A distinguishing feature of PL neurons was the generally low magnitude of their modulations. A Tukey Kramer multiple comparison test confirmed that the magnitude of the modulations differed between all recording sites at $p<0.0001$ with the exception of BL vs. PL ($p=0.1545$).

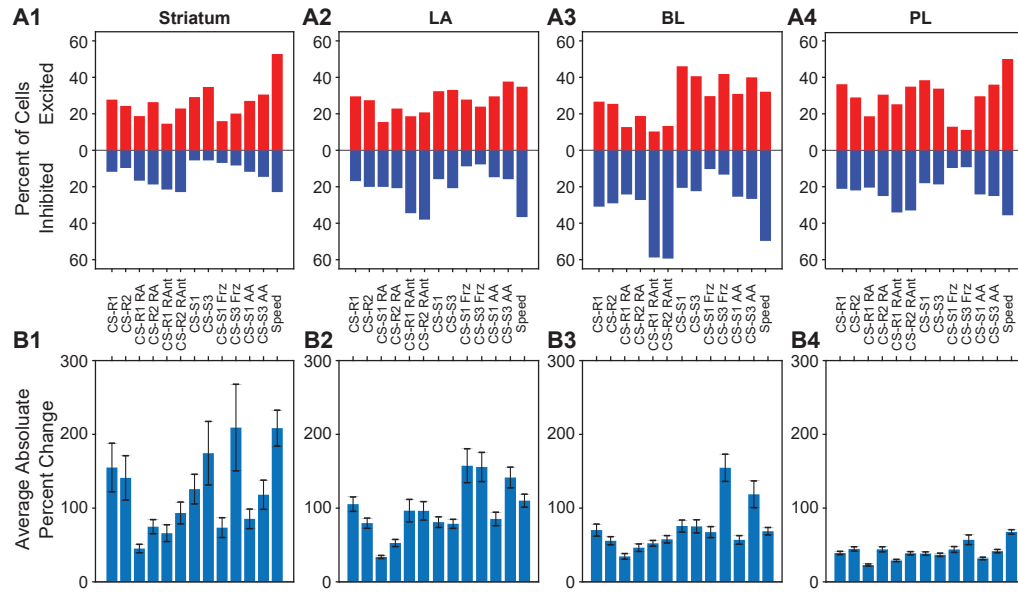


Figure 23. Comparison between coding of RRI task variables by striatal, LA, BL, and PL neurons, as determined using the GLM.

Proportion of presumed PNs (y-axis) in the striatum (A1), LA (A2), BL (A3) and PL (A4) that exhibited excitatory (red) or inhibitory (blue) coding of different task variables used in the GLM (x-axis). (B) Absolute average \pm SEM modulation of the firing rates of striatal (B1), LA (B2), BL (B3) and PL (B4) PNs in relation to each variable used in the GLM. Abbreviations: AA, active avoidance; Frz, freezing; RA, reward approach; RAnt, reward anticipation.

To further characterize coding at the four sites, we correlated the cells' peak modulations by all task variables. As we reported previously (Kyriazi et al., 2018), LA cells tended to show correlated modulations by task events of the same valence (upper left and lower right quadrants in **Fig. 24A2**), and lower or negative correlations between modulations by task events of different valences (lower left and upper right quadrants in **Fig. 24A2**). By contrast, this pattern was less apparent among BL neurons (**Fig. 24A3**) and even less so in PL cells (**Fig. 24A4**). To quantify this aspect, we computed the d-prime score (d' ; Keene et al., 2016; McKenzie et al., 2014), which measures the separation between different coding dimensions based on the distributions of the

correlation coefficients in the similarity matrices. The d' metric was calculated for each coding dimension by comparing the degree to which within-valence correlations exceeded those between-valence (see Methods). Consistent with the impression gained from visual inspection of the matrices, valence coding (**Fig. 25A**) was significantly lower in PL (0.75) than LA (1.66) and BL (1.43; bootstrap resampling, $p < 0.05$) but did not differ from the striatum (0.83). In contrast, the d' metric for active vs. passive behaviors (**Fig. 25B**) was significantly higher in the striatum (1.26) than in LA (0.50) and BL (0.22; bootstrap resampling, $p < 0.05$) but did not differ from PL (0.86; bootstrap resampling, $p > 0.05$).

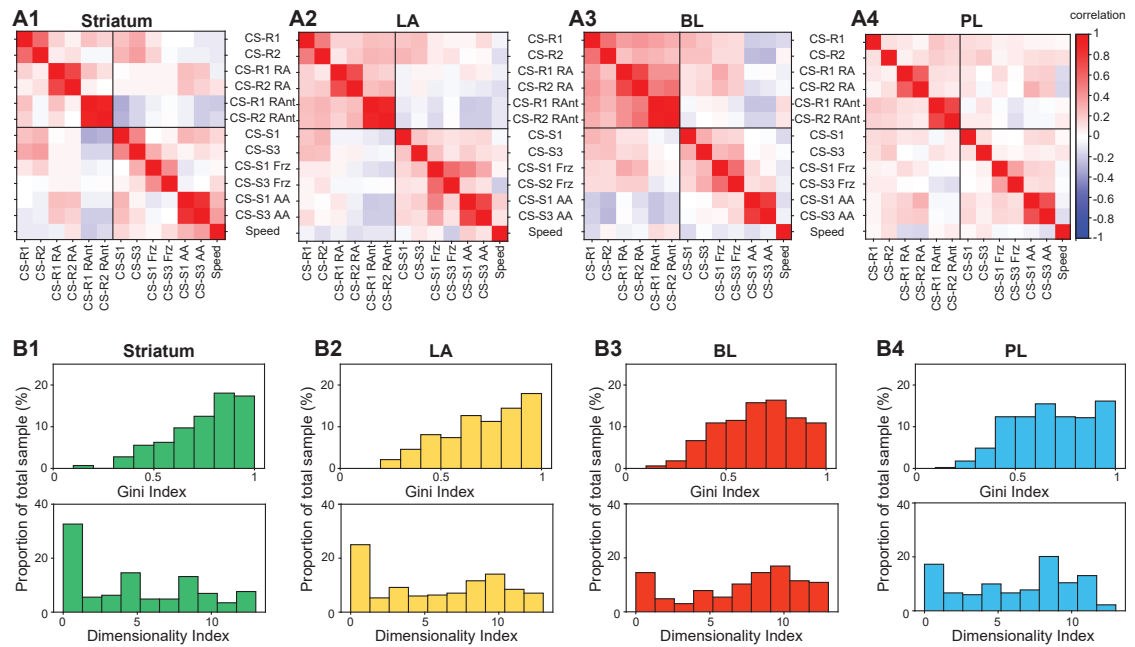


Figure 24. Contrasting representation of task variables in the striatum, amygdala, and prelimbic cortex.

Presumed PNs in the striatum (1), LA (2), BL, (3), and PL (4). (A) Spearman correlation matrices between activity elicited by different task variables. As indicated by the color scale on the left, warmer and cooler colors indicate positive and negative correlations, respectively. (B) Frequency distributions of Gini index. (C) Frequency distributions of dimensionality index. Abbreviations: AA, active avoidance; Frz, freezing; RA, reward approach; RAnt, reward anticipation.

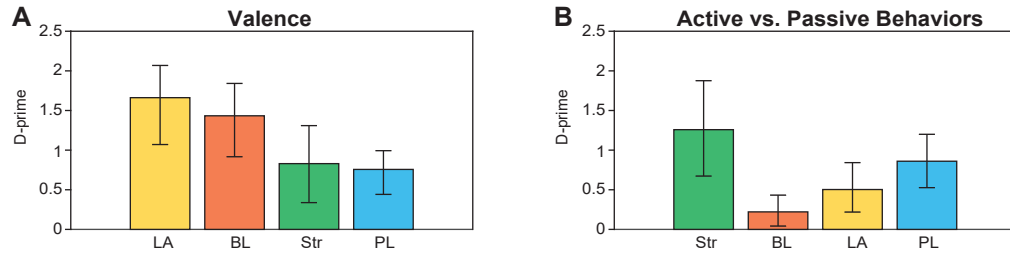


Figure 25. D-prime values.

D-prime values for valence (A) and active vs. passive (B) codes. Error bars represent 95% confidence intervals.

To compare how coding of the various dimensions is allocated within neurons in the striatum, LA, BL, and PL, we computed the Gini index, a measure of statistical dispersion originally used in economics to measure income equality (Gini, 1921). A high Gini index indicates a less distributed code where neurons encode different dimensions unevenly, whereas a low Gini index indicates a more distributed code where neurons encode different dimensions more equally. As shown in **figure 24B** (top), striatal and LA neurons tended to have higher Gini indices than BL and PL neurons. Wilcoxon Ranksum tests confirmed that the Gini indices of LA and striatal neurons did not differ from each other ($p=0.091$) but were significantly higher than in BL (BL vs. LA, $p=0.011$; BL vs. striatum, $p<0.0001$) and PL neurons (PL vs. LA, $p=0.026$; PL vs. striatum, $p<0.0001$), with no difference between PL and BL cells ($p=0.44$). Hence, BL and PL neurons form a more distributed representation of task features than striatal and LA neurons.

Complementing these results, the dimensionality index, which measures how many features each neuron encodes (see Methods) was lowest among striatal and LA neurons and highest among BL and PL cells (**Figure 24B**, bottom). Wilcoxon Ranksum

tests confirmed that the dimensionality indices of striatal neurons were lower than those of LA ($p=0.023$), BL ($p<0.0001$), and PL ($p<0.0001$) neurons. In addition, these tests revealed that the dimensionality indices of LA cells were lower than that of BL cells ($p=0.002$) but did not differ from PL cells ($p=0.217$), with BL cells having higher dimensionality indices than PL neurons ($p<0.007$).

4.5 Population analyses reveal multidimensional representations

Together, the results obtained so far suggest that neurons in the four regions encode information differently. Striatal and LA neurons tend to have strong modulations and encode few task features. By contrast, PL and BL neurons represent multiple features through small fluctuations in their firing rates.

To assess the impact of these differences in the representation of task features at the population level, we performed dimensionality reduction into a 3-dimensional space using principal component analysis (PCA; see General Methods) across the different trial types (CS-R1, CS-R2, CS-S1 and CS-S3; **Fig. 26**). Given that the timing of behaviors varied between trials and that firing rates fluctuated depending on the behaviors emitted, we divided each trial into four equal epochs and time-normalized each epoch by assigning them a fixed number of bins. The four epochs represented the baseline period, the CS-onset period (before any behaviors occurred), the behavior period (RA or AA), and the rest of the trial until US onset.

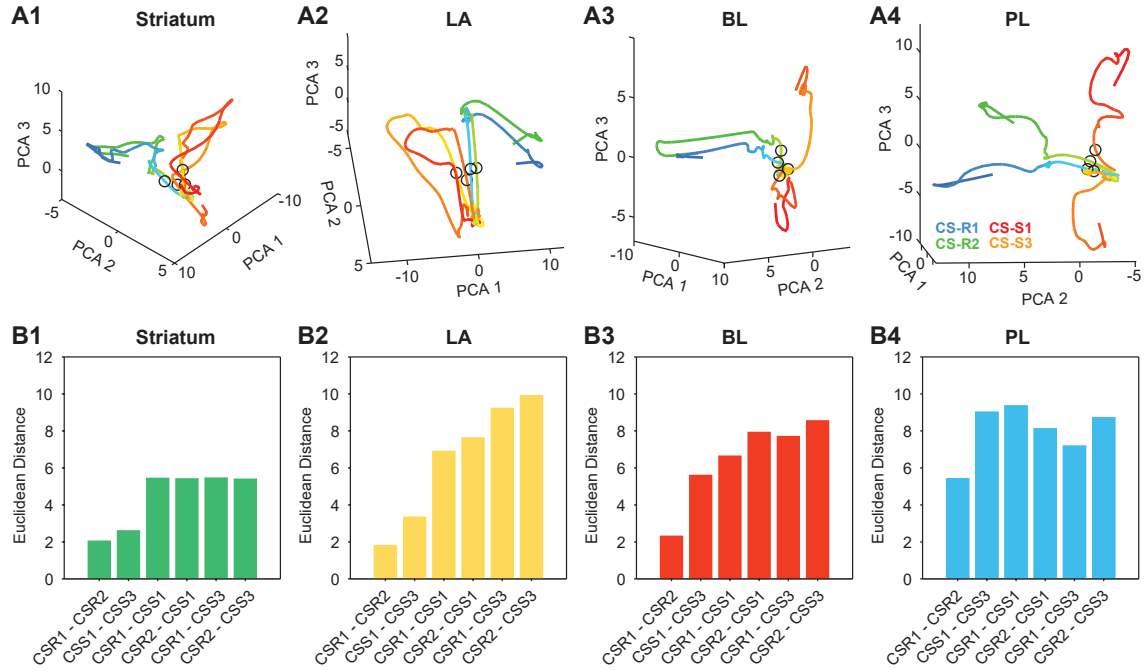


Figure 26. Principal component analysis trajectories by trial type.

(A) Time-normalized trajectories in 3-dimensional PCA space starting 5s before to 10s after CS-onset for STR (A1), LA (A2), BL (A3), and PL (A4) neurons. Colors go from faded to darker across time. The black circle denotes the onset of the stimulus. The 3-dimensional space has been rotated for each structure to show the best dimension of separation. (B) Mean Euclidean distances calculated between each trial type for STR (B1), LA (B2), BL (B3), and PL (B4) neurons.

Figure 26A shows the PCA trajectories for each brain region, starting 5 s before trial onset until 10 s after trial onset (pale to dark colors, respectively). The PCA trajectories revealed striking differences between the four structures. Striatal and LA cells showed trial trajectories separated by valence (Fig. 26A1,2). That is, within CS types (e.g. CS-R1 and CS-R2) trials followed similar paths, whereas between CS types (CS-R vs. CS-S), trial trajectories diverged (Fig. 26A1,2). In PL by contrast, trial paths differed within CS types and by valence (Fig. 26A4). Finally, BL neurons showed an intermediate pattern where trial paths were well separated for only one of the trial types (CS-Ss) while showing divergence as a function of valence (Fig. 26A3). Notably, the divergence

observed in PL trajectories does not appear to be driven by a larger proportion of cells modulated by position information (**Fig. 27**).

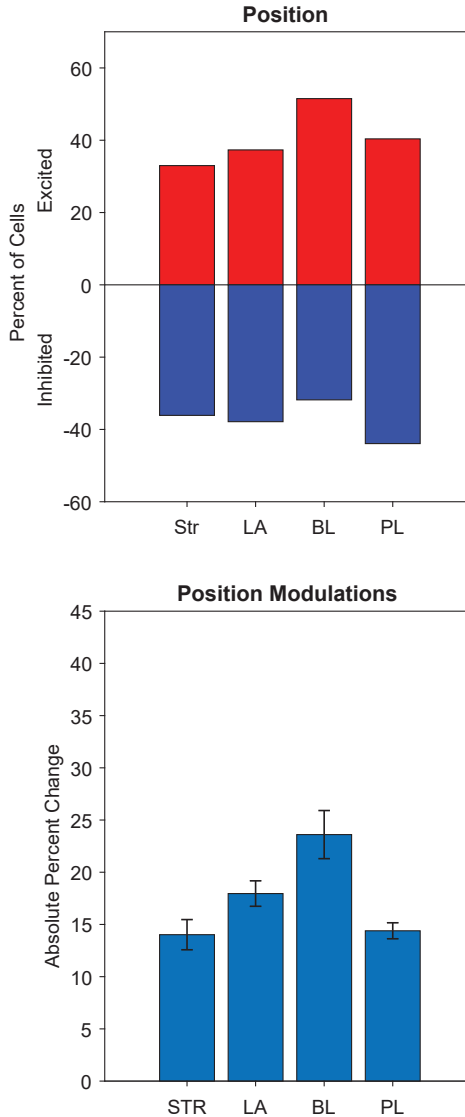


Figure 27. Neural representation of position based on the GLM.

Top: Percent of cells excited (red) and inhibited (blue) by position for presumed PNs recorded in striatum, LA, BL, and PL as estimated by the GLM. Bottom: Absolute average \pm SEM modulation of the firing rates of striatal, LA, BL and PL PNs in relation to position used in the GLM.

To quantify the visual representation of PCA trajectories, we compared the Euclidean distances in 3-dimensional space between trial types (**Fig. 26B**). Euclidean

distances were significantly different between structures ($\chi^2(3)=85.81$, $p<0.0001$), with PL (**Fig. 26B4**) showing higher distances compared to striatum (**Fig. 26B1**), LA (**Fig. 26B2**), and BL (**Fig. 26B3**; Tukey-Kramer post-hoc, all $p<0.01$). Additionally, all other distances differed except for LA and BL ($p=0.9669$).

Since our PCA analyses indicate that population activity diverges based on trial type in PL, then a linear decoder should identify trial type more reliably based on the activity of PL than striatal or BL and LA neurons. Algorithms such as support vector machines implement decoders that are physiologically plausible (Baker, 2003; Poirazi, Brannon, & Mel, 2003). Thus, to test this prediction, we trained a support vector machine with 10-fold cross validation on a resampled population of 150 cells from each brain region and compared decoder performance on individual trial types (**Fig. 28**; see General Methods). Since a binary decoder only allows for two-class comparisons, we trained one decoder on identifying CS-R1 vs. CS-R2 trials and another decoder on CS-S1 vs. CS-S3 trials. We then pooled the performance accuracy of the two decoders to calculate a performance score for trial identity. Indeed, the decoders were superior at identifying trial type based on the activity of PL than BLA or striatal neurons, during both, the stimulus (**Fig. 28A1**) and behavior (**Fig. 28B1**) epochs. Consistent with the higher response selectivity we observed earlier in PL using PSTHs (**Figs. 17 and 20**), CS identity was decoded with higher accuracy by PL neurons (93%) than BL (70%), LA (68%), and striatal (55%) neurons (**Fig. 28A1**). While decoding accuracy during the behavior epoch was high with all cell types (striatal: 74%, LA: 84%, BL: 92%), PL neurons still outperformed the others with 98% accuracy (**Fig. 28B1**). Nevertheless, all decoders performed above chance as calculated by shuffling trial labels (dashed line).

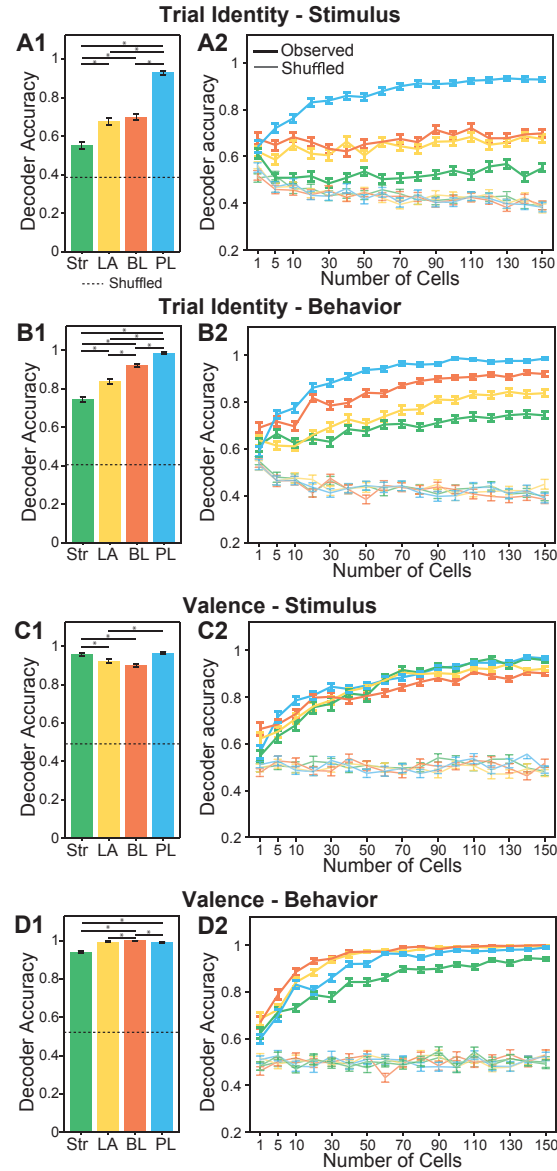


Figure 28. Trial type and valence decoders.

(A1) Decoder accuracy of trial type (CS-R1 vs. CS-R2 and CS-S1 vs. CS-S3) during the stimulus period before behaviors are initiated. Dashed line represents shuffled trial type, indicating chance. Error bars represent the SEM based on 50 decoder repetitions. (B1) Decoder accuracy of trial type during active CRs (CS-R1 RA vs. CS-R2 RA and CS-S1 AA vs. CS-S3 AA). (C1) Decoder accuracy of valence (CS-Rs vs. CS-Ss) during the stimulus period before behaviors are initiated. (D1) Decoder accuracy of valence active CRs (RA vs. AA). (A2-D2) Decoder accuracy as population size increases. Dark colored lines represent observed decoder performance. Light colored lines represent decoder performance with shuffled trial labels.

Next, we tested whether valence could be decoded from the same population activity. Accordingly, we collapsed CS-R1 and CS-R2 trials into a CS-R category as well as CS-S1 and CS-S3 trials into a CS-S category and then trained the decoder to distinguish between CS-R vs. CS-S trials. Decoding accuracy for valence was above 90% in all brain regions for both stimulus (**Fig. 28C1**) and behavior (**Fig. 28D1**; AA vs. RA behaviors) periods.

To determine how the size of the population influences decoding accuracy, we increased the number of cells (from 1 to 5, and then 10 to 150 in steps of 10 units). This approach revealed significant differences between structures for trial identity during the stimulus (**Fig. 28A2**; $\chi^2(3)=54.46$, $p<0.0001$) and behavior (**Fig. 28B2**; $\chi^2(3)=34.29$, $p<0.0001$) periods as well as for valence during behavior (**Fig. 28D2**; $\chi^2(3)=18.11$, $p<0.001$). Decoding trial identity during the stimulus period resulted in improved performance with growing population size only in PL. In BLA, the number of cells did not contribute to the decoder's performance, indicating that few cells hold a weak representation of trial identity information at the stimulus level. Trial identity performance decreased as the number of striatal cells was increased but remained above chance. This early decrease could be an artifact of the few number of trials used. There were only 10 trials total used in this analysis (5 CS-R, 5 CS-S) meaning that 10-fold cross validation would use only 9 trials to calculate performance, leading to chance levels $<50\%$, hence the decrease below chance when few cells are used to decode trial identity. A Tukey-Kramer post hoc test showed that all structures differed in decoding performance for trial identity during the stimulus period ($p<0.05$), except LA and BL ($p=0.5345$). During the behavior period, trial identity decoder accuracy increased in all

brain structures with increasing population size, but only striatum vs. BL ($p=0.001$) and PL ($p<0.0001$), as well as LA vs. PL ($p<0.0001$) were significantly different from each other. Finally, valence decoding during the stimulus period did not differ between structures (**Fig. 28C2**; $\chi^2(3)=4.05$, $p=0.2557$), while valence during the behavior period was lower for striatum compared to LA ($p=0.0048$) and BL ($p<0.0001$).

The prevalence of the valence code across regions was surprising and raises the question of whether the valence code in LA and BL is similar to the one in PL and striatum. One way to address this is to determine whether the coding of valence during the CS period generalizes to the CR period. Cross-temporal generalization analyses capture this by training a decoder during one trial period, and testing during another period. Valence codes in LA and BL generalized better between CS and CR periods than those in PL and striatum (**Fig. 29A**). These analyses also showed that the PL code for trial identity generalized between trial periods, suggesting that its ensemble activity codes for trial identity in a manner that is consistent across the entire trial (**Fig. 29B**).

Together the decoder results indicate that PL neurons are better at holding information about multiple representations simultaneously including trial identity and valence, while BLA and striatal units are more specialized for valence coding. Additionally, increasing population size influences decoding accuracy differently depending on the brain structure and the item being decoded.

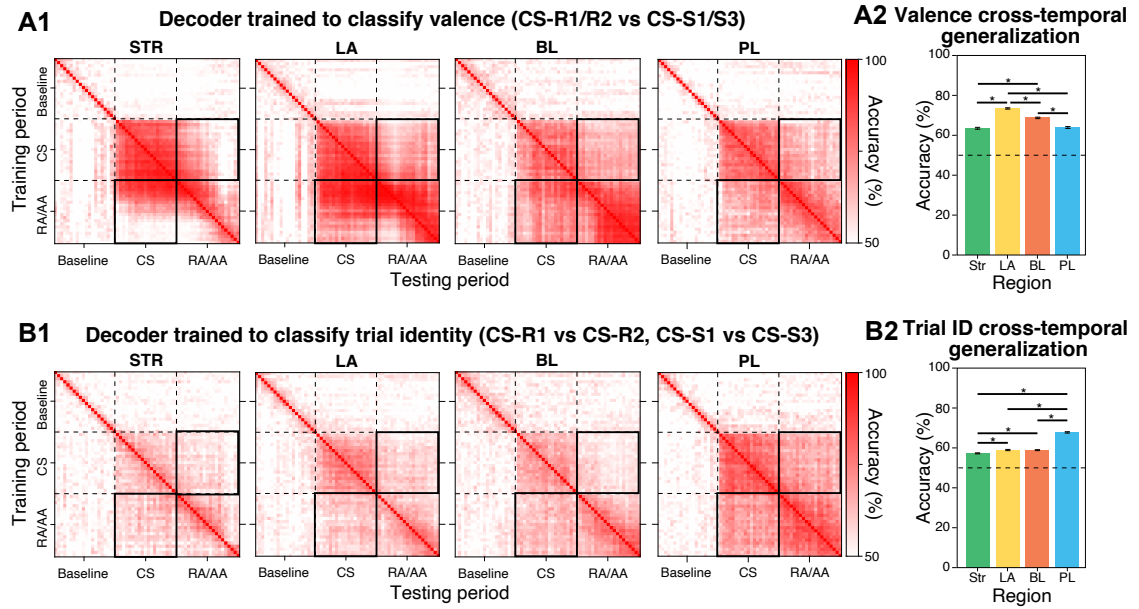


Figure 29. Cross-temporal generalization of the SVM decoder.

Cross-temporal generalization of the SVM decoder. **(A1)** SVM decoders were trained for valence on one time point during the trial and tested on another. Training ensembles were comprised of 150 PNs. Dashed lines indicate trial periods, solid squares encompass train/test pairs that were averaged together to measure cross-temporal generalization. **(A2)** The mean response during the cross-temporal generalization periods were recalculated 50 times using resamples of trials and units from our database. The distribution of mean differences between regions was bootstrap recalculated 10,000 times and deemed significant when 0 difference was either below 2.5% or above 97.5% of the difference distribution. LA showed the strongest generalization of the valence code between stimulus and behaviors periods, followed by BL, PL and striatum. **(B1)** and **(B2)** are similar to **A1** and **A2**, but with the decoder trained on trial identity.

4.6 Conclusions

In keeping with their similar connectivity, the BLA and mPFC are involved in closely related functions, raising the question of what is the computational advantage of having both. Here, I tested the hypothesis that mPFC and BLA neurons code information differently, potentially allowing them to implement distinct types of input-output functions. Accordingly, I compared the coding properties of PL, LA, and BL neurons during the performance of a task where rats generate different defensive or appetitive

behaviors depending on CS location on each trial. The findings indicate that while amygdala and mPFC neurons have specialized functions, they also form overlapping representations of task features. Hence, the amygdalo-prefrontal network is not characterized by clear-cut regional differences in the identity of the variables encoded but by gradients in the representation of the same variables and how they are encoded.

Chapter V:

General Discussion

5.1 Independent encoding of CSs and CRs by BLA neurons

In chapter III, several lines of evidence indicated that sensory responses and CRs are encoded concurrently, yet independently, by the same BLA neurons. For instance, separately plotting BLA firing rates when CRs were emitted at different latencies revealed an early CS-locked response and a later component whose latency increased with that of the CRs. Similarly, comparing BLA activity on correct and error trials provided evidence for two response components: an early sensory-related part that was present on correct and error trials and a late component, present on correct trials but nearly absent on error trials.

The fluctuating character of the late, CR-related component despite the stability of the early CS-locked phase is inconsistent with the view that, due to the potentiation of CS inputs on BLA neurons, CRs are automatically triggered by CSs. Instead, these results suggest that the receipt of sensory information about CSs and the emission of CRs are distinct, yet related processes. Most likely, interactions between the BLA and other structures such as the medial prefrontal and temporal cortices (Pitkänen, 2000) regulate whether and when a given CS will elicit a CR of a particular type. Such interactions

would account for the behavioral flexibility exhibited by animals depending on context (Perusini & Fanselow, 2015).

The realization that BLA neurons separately, yet concurrently encode CSs and CRs sheds new light on earlier findings. For instance, using Ca^{2+} -imaging of large groups of BLA neurons in vivo, Grewe et al. (2017) reported that after fear conditioning, the representation of the CS by BLA neurons morphed into that of the US. Since our results indicate that individual BLA cells show similar activity in relation to behaviors of the same valence, and considering that after (but not before) training, both USs and CSs elicited negatively valenced behaviors, the change in representation Grewe et al. (2017) observed could have resulted from the shift in the behavioral outcome of the CS.

5.2 At the level of individual cells, valence-coding is not prevalent in the BLA

Here, valence-coding neurons were defined as cells that showed a firing rate increase in relation to at least one CS and CR of the same valence coupled to an inhibition or no change in activity in relation to all CSs and CRs of the opposite valence. Under this definition, few valence-coding cells were observed in the BLA (<20%). Given the proportions of cells whose activity was modulated by the various CSs and CRs, the incidence of valence-coding cells did not exceed chance. Yet, the GLM analyses indicated that valence information is encoded in the collective activity of BLA neurons. To gain insights into this apparent contradiction, the estimated modulation of BLA activity by CSs and CRs was correlated. This revealed that in LA, the firing rate modulations by CSs of opposite valence were positively correlated whereas those

associated with CRs of opposite valence were not. Thus in this paradigm, valence information is mostly derived from the modulation of activity by CRs.

Of note, the correlated modulations of BLA activity by the CS-Rs and CS-Ss were expected in this paradigm given that rats had to rely on location to determine whether similar light stimuli were CS-Ss or CS-Rs. The fact that BLA neurons showed differentiated modulations to CRs of opposite valence despite the similar sensory properties of the CSs casts further doubt on the notion that potentiated CS inputs onto distinct subsets of valence-coding BLA neurons is the sole factor supporting the differentiated expression of conditioned emotional behaviors. Rather, these findings again suggest that in between the receipt of sensory information about CSs by BLA neurons and the emission of CRs are additional processing steps where other streams of information, in this case about place, determine the outcome of the CS.

Another unexpected finding in this study was the positive correlation, particularly marked in LA, between the activity modulations associated with freezing and active avoidance. Although these CRs constitute markedly different defense strategies, BLA cells, even valence-coding cells showed similar activity modulations in relation to both behaviors. This observation suggests that while different subsets of BLA neurons may generate distinct CRs via their particular outputs, it is incumbent on target effector neurons to decode what CRs should be generated. Presumably, other inputs allow them to disambiguate the significance of BLA inputs at any given time.

5.3 Multidimensional coding by BLA neurons

While explicit coding of valence by individual neurons was not prominent in the BLA, valence information was present at the population level. In fact, MDS analyses suggested that several coding dimensions are represented in the ensemble activity of BLA neurons, including valence, speed, and active vs. passive behaviors. Unexpectedly, valence coding was primarily seen in the behavioral domain. The preferential association of valence coding with behavior rather than CSs is reminiscent of a trace conditioning study in monkeys where valence coding by BLA neurons was more prominent during the trace period, shortly before the CR (Paton, Belova, Morrison, & Salzman, 2006). Furthermore, the active-passive coding dimension was orthogonal to valence coding, which suggests that LA and BL neurons can concomitantly represent different types of information with minimal interference. Interestingly, the active-passive and speed dimensions had opposite directions in LA vs. BL. Whereas in LA, a negative modulation by speed was associated with a positive modulation by valence-behavior, the opposite relationship was found in BL. The origin of these differences remains unclear.

Consistent with these findings, other behavioral paradigms have uncovered coexisting codes in BLA ensemble activity. For instance, Munuera et al. (2018) reported that in non-human primates, cells that encode the reward value of non-social stimuli also encode the hierarchical rank of conspecifics. Moreover, in a paradigm where the reward associated with different stimuli varied in a context-dependent fashion, BLA cells encoded context, CS identity, and reinforcement expectations (A. Saez et al., 2015).

Overall, these considerations suggest that conditioned emotional behaviors do not depend on the recruitment of discrete, dedicated subsets of cells that explicitly encode a

particular valence, but on multiplexed representations distributed across the ensemble activity of BLA neurons. These high-dimensional coding schemes confer major computational advantages (Fusi, Miller, & Rigotti, 2016), notably a dramatic increase in the range of input-output relations that can be mapped by the same network. While this perspective is relatively novel for the amygdala, its ubiquity in the brain has long been recognized. For instance, high-dimensional population coding has been documented in many cortical areas (Fusi et al., 2016; McKenzie et al., 2014; Morris, Bremmer, & Krekelberg, 2016; Rigotti et al., 2013). Even the control of spinal motor neurons by cortical and rubral cells depends on such distributed codes (Fetz, 1992).

The presence of distributed coding would seem to contradict the numerous demonstrations of emotional behaviors controlled by specific cell types in the amygdala. While our findings argue against a strong view where neurons are treated as labeled lines for specific emotional behaviors and states, they are consistent with the fact that those variables are only partially correlated with anatomical position (Beyeler et al., 2018), projection target (Beyeler et al., 2016) or genetic profile (Kim, Pignatelli, Xu, Itoharu, & Tonegawa, 2016). Thus, it might be better to view coding in the amygdala as a spectrum, not a switchboard.

5.4 Comparing BLA and mPFC activity

In keeping with their similar connectivity and responsiveness to task events (i.e. stimuli, behaviors), BLA and mPFC neurons are involved in closely related functions, raising the question of what is the computational advantage of having both regions. In chapter IV, I tested the hypothesis that mPFC and BLA neurons code information

differently, potentially allowing them to represent distinct task dimensions (e.g. valence). Accordingly, I compared the responses of PL, LA, and BL neurons during a task where rats generated different defensive or appetitive behaviors depending on CS location. The findings showed that while specific task events were represented by individual neurons in PL, LA, and BL, cells differed in the combinations of events they responded to. At the population level, these differences scaled-up to variations in their proclivity to encode certain task dimensions. Hence, the amygdalo-prefrontal network is not characterized by clear-cut regional differences in the identity of the task features encoded, but by gradations in the representation of task dimensions.

5.5 Gradations in task representations

On the surface, the task-related activity of PL neurons appeared similar to that of LA and BL cells. For instance, a nearly identical proportion of neurons at the three sites significantly increased their firing rate in response to at least one CS. Also, cells with CR-related activity seemed as prevalent in PL, LA, and BL. Moreover, at the three sites, CS-evoked activity showed a strong dependence on behavior, with the same CS eliciting larger responses on trials where rats emitted the appropriate CR than on error trials. Beneath these superficial similarities however, there were conspicuous differences. First, the cells' responses to different CSs of the same valence were more dissimilar in PL than in LA or BL. That is, many more PL than LA or BL cells had inhibitory responses to only one CS of a given valence. PL neurons also showed greater selectivity for CRs evoked by different CSs: whereas in LA and BL, most CR-activated cells were recruited similarly irrespective of the CS being presented, the CR-related activity of most PL neurons varied

depending on the CS that evoked it. This encoding of specific conjunctions of task features is known as mixed selectivity, which has been described previously in other prefrontal areas including lateral PFC (Lindsay, Rigotti, Warden, Miller, & Fusi, 2017; Rigotti et al., 2013), orbitofrontal, and anterior cingulate cortices (A. Saez et al., 2015). Furthermore, mixed selectivity is thought to provide a computational advantage for multidimensional representations (Fusi et al., 2016).

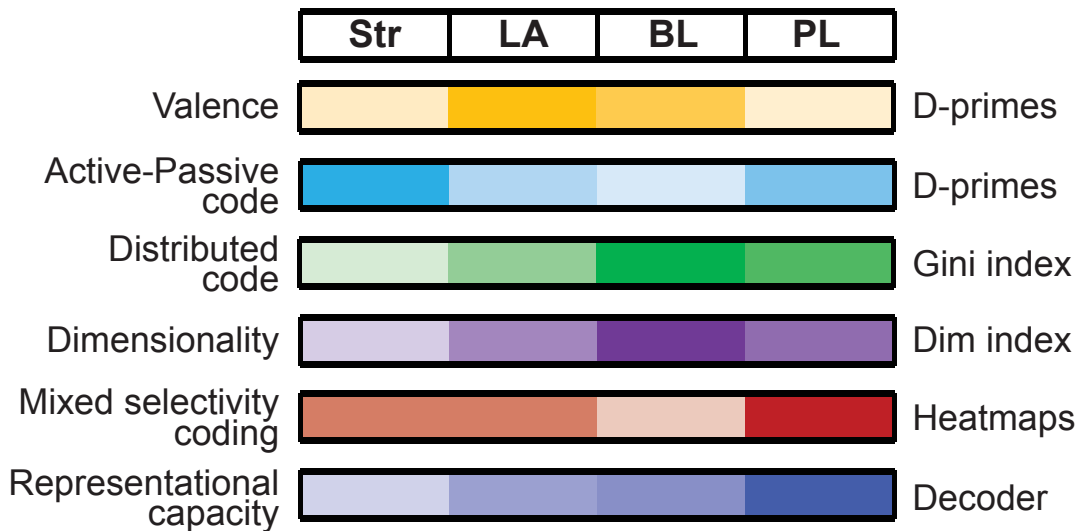


Figure 30. Representation gradients across the amygdalo-striatal-prefrontal network.

Gradient scales of representation for items listed on the left side. Scale values for each brain region were calculated based on measures listed on the right side.

That LA, BL, and mPFC neurons encode overlapping task features while showing differences in how they are represented at the population level, suggests that these regions serve partially distinct functions. Indeed, directly comparing these population representations between regions does not reveal discrete differences in what task dimensions are encoded, but instead differing propensities for the representation of some

dimensions over others, hereafter referred to as representation gradients. In our study, we observed several instances of this phenomenon (**Fig. 30**). First, we observed a gradient of valence coding across the amygdalo-striatal-prefrontal network. While it is well known that LA and BL cells are modulated by valence (Belova et al., 2008; Kyriazi et al., 2018; Namburi et al., 2015; Paton et al., 2006; Sangha et al., 2013), striatal and PL neurons were found to represent valence to a lesser degree, as evidenced by their low d' index. Nevertheless, valence could be easily decoded from the population activity of all four regions, as indicated by the linear decoder results (**Fig. 28A-D**). However, the ability for a decoder to derive valence from striatum and PL does not necessarily mean that those areas code for valence in the same way as LA and BL. In LA and BL, neurons tended to respond similarly to all CSs and CRs with the same valence (**Figs. 24A, 25A**), while this was not the case for either striatum or PL. An additional confirmation of this came from cross-temporal generalization analyses of our decoders (**Fig. 30**), which found that striatal and PL decoders trained on the CS period did not perform well during the behavior period. Thus, while the valence code in LA and BL strongly generalizes across conditions, in striatum and PL the generalization is weaker.

Another gradient of representation was found for active vs. passive behaviors. In this case, striatal cells had the strongest representation, followed by PL, and finally amygdala cells (**Fig. 30**). That striatal cells can easily distinguish different behavioral strategies is consistent with the involvement of the striatum in motor functions (Wall, De La Parra, Callaway, & Kreitzer, 2013). Similarly, mPFC neurons respond differently to movement excitation and inhibition during fear conditioning (Halladay & Blair, 2015), and an inhibitory signal in PL was found to mediate active avoidance (Diehl et al., 2018).

Finally, given that amygdala neurons weakly distinguish between active and passive strategies, it is unlikely that they orchestrate the execution of such behaviors, but rather integrate inputs from mPFC and provide a bias for approaching or avoiding stimuli based on valence (Lazaro-Munoz, LeDoux, & Cain, 2010; Sangha, Diehl, Bergstrom, & Drew, 2020).

5.6 Gradations in the manner of population representations

Besides *what* is encoded by BLA and PL neurons, there is also the question of *how* the codes are represented by these cell populations. Here too, there is a continuum in the form of representation across these regions (**Fig. 30**), as quantified with the Gini, dimensionality indices, and event-evoked firing rates. While individual PL and BL neurons tended to weakly respond to many different task events, LA and striatal cells coded fewer task features with stronger changes in their firing rates. Although it may seem odd that PL was more similar to BL than the neighboring LA nucleus in this respect, this finding is consistent with the connections between these three structures. That is, PL forms stronger reciprocal connections with BL than LA (Krettek & Price, 1977; McDonald, 1991; McDonald et al., 1996). Moreover, BL projections to LA are sparse (Krettek & Price, 1978; Pitkanen et al., 1997).

Another instance of graded regional differences in coding was found in their representational capacity (**Fig. 30**). Using PCA analyses, we observed that whereas striatal and LA populations showed trial trajectories only separated by valence (or its associated CSs and CRs), PL exhibited a unique trajectory for each trial type. Between these extremes, BL neurons distinguished valence, but they also differentiated the two

types of CS-S trials. The ability of PL neurons to maintain independent representations of each trial type indicates that the population activity of PL neurons holds more information than that of BLA or striatal neurons. The more distinct representations found in PL may prevent interference when learning new associations (Bartolo, Saunders, Mitz, & Averbeck, 2019).

If these representations are physiologically meaningful, then they must be decodable by downstream neurons. Indeed, decoders were superior at identifying trial identity based on the activity of PL than BLA or striatal neurons, during either the stimulus or behavior periods (**Fig. 28A-D**). Moreover, gradually increasing the size of the cell populations used in these analyses for the stimulus period only improved decoder performance with PL neurons, suggesting that although BLA and striatal cells hold some information about trial identity, redundancy between cells limits population level decoding of trial types. This may arise because BLA neurons respond similarly to stimuli and behaviors of the same valence, in effect causing them to act as ‘valence-detectors.’ Indeed, smaller sized BLA ensembles outperformed those from PL and striatum for encoding valence (**Fig. 28D**). But as population size increased, PL caught up, achieving comparable performance. These results indicate that even though PL neurons show small task-related fluctuations in firing rates, their population activity has greater representational capacity, as it can encode both trial identity and valence information simultaneously. This is in line with recent observations that the activity of prefrontal neurons forms a geometric pattern in multidimensional space that makes it possible to decode abstract representations such as context and history of reward value (Bernardi et al., 2018).

5.7 Codes and computations

Different situations often call for distinct behavioral strategies, and correspondingly different computations (Headley et al., 2019). The activity patterns that accompany these processes will reflect these differences and constrain the kinds of information that downstream areas can extract. Our results indicate that PL neurons encode a wide variety of contingencies, that LA and striatum preferentially represent valence and behaviors, while BL lies between PL and LA. Prefrontal areas such as PL are important for planning and executing context-dependent behaviors (Hyman, Ma, Balaguer-Ballester, Durstewitz, & Seamans, 2012; Ma et al., 2016; Moorman & Aston-Jones, 2015), while LA links stimuli with their value (Repa et al., 2001; Tye et al., 2008). This gradient, from the generic (PL) to the specific (LA), may reflect the distinct cognitive demands placed on these regions (Mobbs, Headley, Ding, & Dayan, 2020). And yet, these areas are coactive during even the simplest behaviors (e.g. Vetere et al., 2017). Thus, the gradients of representations we observed across the amygdalo-prefrontal network may enable the sharing of information, and provide parallel portraits of task contingencies, allowing for flexibility and redundancy in the production of emotional behaviors.

5.8 Gaps and future directions

While our study helped reveal multidimensional representations in the amygdalo-prefrontal network, it suffers from some gaps and shortcomings that could be addressed in future research. The main drawback of our task is the unique timing of each trial:

animals may perform any of the behaviors at different timepoints across the trial and for different durations of time. While this lack of a fixed structure makes the task more naturalistic, it complicates analyses. For this reason, many of the analytical techniques I used relied on the peak of the modulation of cells responses (e.g., GLM, correlation matrices, and modulations of firing rate, MDS) regardless of the timing in the trial. Some attempts to overcome this were made by normalizing the timing of each trial as the analyses performed for the PCA (**Fig. 26**), SVM (**Fig. 28**), and cross-temporal generalization (**Fig. 29**). However, normalizing time in a trial results in stretching or shortening of trial periods leading to slight alterations in firing rate. This shortcoming may have overstated some of the multidimensional representations found through the MDS analysis. Examples of studies with fixed trial types come from the monkey literature (Belova et al., 2008; Paton et al., 2006; A. Saez et al., 2015). Future research should find ways to standardize trials and use analyses to understand the temporal dynamics of cell activity across trials.

Another limitation of this study was our inability to analyze the US responses of the cells due to artifacts. Studies have indicated that after conditioning, CS responses become more similar to US representations (Grewe et al., 2017). This was not something we could study in our task due to electrical artifacts from footshocks and licking. However, it is an important question and one that will contribute to our understanding of amygdala and prefrontal representations. Hence future studies should use techniques such as calcium imaging to overcome electrical artifacts and study population representations of US responses.

Finally, even though we show that cells fire in relation to different conditioned behaviors, it is unclear whether these cells are mediating these actions. Most likely their contribution is not one of behavior execution but rather a bias signal guiding the animal towards a certain brain state. As a result, the animal may be more likely to engage in behaviors of a certain valence such as reward approach or anticipation and avoidance or freezing. As other regions provide more refined information, the bias for a certain behavior may go beyond valence. We observed this signal in PL, where cells did not only signal the valence of a given behavior but differentiated when the behavior was performed in response to each CS-R or CS-S.

The coordination of different brain regions giving similar signals at the same time, most likely result in guiding downstream effectors such as PAG to engage in specific behaviors. Indeed, some studies have indicated that signals leading to PAG activation drive different defensive behaviors, such as escape, freezing, and avoidance (Evans et al., 2018; Fadok et al., 2017; Rozeske et al., 2018). Future studies using the RRI task should try to understand how the signals from BLA and PL drive downstream effectors, including PAG, to produce different appetitive and aversive behaviors.

List of references:

- Allen, G. V., Saper, C. B., Hurley, K. M., & Cechetto, D. F. (1991). Organization of visceral and limbic connections in the insular cortex of the rat. *Journal of Comparative Neurology*, 311(1), 1-16.
- Ambroggi, F., Ishikawa, A., Fields, H. L., & Nicola, S. M. (2008). Basolateral amygdala neurons facilitate reward-seeking behavior by exciting nucleus accumbens neurons. *Neuron*, 59(4), 648-661.
- Amir, A., Headley, D. B., Lee, S. C., Haufler, D., & Pare, D. (2018). Vigilance-Associated Gamma Oscillations Coordinate the Ensemble Activity of Basolateral Amygdala Neurons. *Neuron*, 97(3), 656-669.e657.
- Amir, A., Lee, S. C., Headley, D. B., Herzallah, M. M., & Pare, D. (2015). Amygdala Signaling during Foraging in a Hazardous Environment. *Journal of Neuroscience*, 35(38), 12994-13005.
- Amorapanth, P., Nader, K., & LeDoux, J. E. (1999). Lesions of periaqueductal gray dissociate-conditioned freezing from conditioned suppression behavior in rats. *Learning and Memory*, 6(5), 491-499.
- Baker, J. L. (2003). Is There a Support Vector Machine Hiding in the Dentate Gyrus? *Neurocomputing*, 52-54, 199-207.
- Bartho, P., Hirase, H., Monconduit, L., Zugaro, M., Harris, K. D., & Buzsaki, G. (2004). Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *Journal of Neurophysiology*, 92(1), 600-608.
- Bartolo, R., Saunders, R. C., Mitz, A., & Averbeck, B. B. (2019). Dimensionality, information and learning in prefrontal cortex. *bioRxiv*.
- Beck, K. D., Jiao, X., Smith, I. M., Myers, C. E., Pang, K. C., & Servatius, R. J. (2014). ITI-Signals and Prelimbic Cortex Facilitate Avoidance Acquisition and Reduce Avoidance Latencies, Respectively, in Male WKY Rats. *Frontiers in Behavioral Neuroscience*, 8, 403.
- Belova, M. A., Paton, J. J., & Salzman, C. D. (2008). Moment-to-moment tracking of state value in the amygdala. *Journal of Neuroscience*, 28(40), 10023-10030.

- Berke, J. D. (2008). Uncoordinated firing rate changes of striatal fast-spiking interneurons during behavioral task performance. *Journal of Neuroscience*, 28(40), 10075-10080.
- Bernardi, S., Benna, M. K., Rigotti, M., Munuera, J., Fusi, S., & Salzman, C. D. (2018). The geometry of abstraction in hippocampus and prefrontal cortex. *bioRxiv*.
- Beyeler, A., Chang, C. J., Silvestre, M., Leveque, C., Namburi, P., Wildes, C. P., & Tye, K. M. (2018). Organization of Valence-Encoding and Projection-Defined Neurons in the Basolateral Amygdala. *Cell Reports*, 22(4), 905-918.
- Beyeler, A., Namburi, P., Gloomer, G. F., Simonnet, C., Calhoon, G. G., Conyers, G. F., . . . Tye, K. M. (2016). Divergent Routing of Positive and Negative Information from the Amygdala during Memory Retrieval. *Neuron*, 90(2), 348-361.
- Bienvenu, T. C., Busti, D., Magill, P. J., Ferraguti, F., & Capogna, M. (2012). Cell-type-specific recruitment of amygdala interneurons to hippocampal theta rhythm and noxious stimuli in vivo. *Neuron*, 74(6), 1059-1074.
- Bittencourt, A. S., Carobrez, A. P., Zamprogno, L. P., Tufik, S., & Schenberg, L. C. (2004). Organization of single components of defensive behaviors within distinct columns of periaqueductal gray matter of the rat: role of N-methyl-D-aspartic acid glutamate receptors. *Neuroscience*, 125(1), 71-89.
- Blanchard, D. C., Griebel, G., Pobbe, R., & Blanchard, R. J. (2011). Risk assessment as an evolved threat detection and analysis process. *Neuroscience and Biobehavioral Reviews*, 35(4), 991-998.
- Blanchard, D. C., Hynd, A. L., Minke, K. A., Minemoto, T., & Blanchard, R. J. (2001). Human defensive behaviors to threat scenarios show parallels to fear- and anxiety-related defense patterns of non-human mammals. *Neuroscience and Biobehavioral Reviews*, 25(7-8), 761-770.
- Borg, I., & Groenen, P. J. F. (2005). Modern Multidimensional Scaling: Theory and Applications. (New York: Springer Science+Business Media, Inc).
- Bravo-Rivera, C., Roman-Ortiz, C., Brignoni-Perez, E., Sotres-Bayon, F., & Quirk, G. J. (2014). Neural structures mediating expression and extinction of platform-mediated avoidance. *Journal of Neuroscience*, 34(29), 9736-9742.
- Bravo-Rivera, C., Roman-Ortiz, C., Montesinos-Cartagena, M., & Quirk, G. J. (2015). Persistent active avoidance correlates with activity in prelimbic cortex and ventral striatum. *Frontiers in Behavioral Neuroscience*, 9, 184.

- Breheny, P., & Huang, J. (2015). Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Stat Comput*, 25(2), 173-187.
- Burgos-Robles, A., Bravo-Rivera, H., & Quirk, G. J. (2013). Prelimbic and infralimbic neurons signal distinct aspects of appetitive instrumental behavior. *PloS One*, 8(2), e57575.
- Burgos-Robles, A., Kimchi, E. Y., Izadmehr, E. M., Porzenheim, M. J., Ramos-Guasp, W. A., Nieh, E. H., . . . Tye, K. M. (2017). Amygdala inputs to prefrontal cortex guide behavior amid conflicting cues of reward and punishment. *Nature Neuroscience*, 20(6), 824-835.
- Burgos-Robles, A., Vidal-Gonzalez, I., & Quirk, G. J. (2009). Sustained conditioned responses in prelimbic prefrontal neurons are correlated with fear expression and extinction failure. *Journal of Neuroscience*, 29(26), 8474-8482.
- Ciocchi, S., Herry, C., Grenier, F., Wolff, S. B., Letzkus, J. J., Vlachos, I., . . . Luthi, A. (2010). Encoding of conditioned fear in central amygdala inhibitory circuits. *Nature*, 468(7321), 277-282.
- Corcoran, K. A., & Quirk, G. J. (2007). Activity in prelimbic cortex is necessary for the expression of learned, but not innate, fears. *Journal of Neuroscience*, 27(4), 840-844.
- Di Lorenzo, P. M., Chen, J. Y., & Victor, J. D. (2009). Quality time: representation of a multidimensional sensory domain through temporal coding. *Journal of Neuroscience*, 29(29), 9227-9238.
- Diehl, M. M., Bravo-Rivera, C., Rodriguez-Romaguera, J., Pagan-Rivera, P. A., Burgos-Robles, A., Roman-Ortiz, C., & Quirk, G. J. (2018). Active avoidance requires inhibitory signaling in the rodent prelimbic prefrontal cortex. *Elife*, 7.
- Doron, N. N., & Ledoux, J. E. (1999). Organization of projections to the lateral amygdala from auditory and visual areas of the thalamus in the rat. *Journal of Comparative Neurology*, 412(3), 383-409.
- Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron*, 76(6), 1057-1070.
- Evans, D. A., Stempel, A. V., Vale, R., Ruehle, S., Lefler, Y., & Branco, T. (2018). A synaptic threshold mechanism for computing escape decisions. *Nature*, 558(7711), 590-594.

- Fadok, J. P., Krabbe, S., Markovic, M., Courtin, J., Xu, C., Massi, L., . . . Lüthi, A. (2017). A competitive inhibitory circuit for selection of active and passive fear responses. *Nature*, 542(7639), 96-100.
- Fanselow, M. S. (1994). Neural organization of the defensive behavior system responsible for fear. *Psychon Bull Rev*, 1(4), 429-438.
- Fanselow, M. S., & Poulos, A. M. (2005). The neuroscience of mammalian associative learning. *Annual Review of Psychology*, 56, 207-234.
- Fetz, E. E. (1992). Are movement parameters recognizably coded in the activity of single neurons? *Behavioral and Brain Sciences*, 15(4), 679-690.
- Floyd, N. S., Price, J. L., Ferry, A. T., Keay, K. A., & Bandler, R. (2000). Orbitomedial prefrontal cortical projections to distinct longitudinal columns of the periaqueductal gray in the rat. *Journal of Comparative Neurology*, 422(4), 556-578.
- Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. *Current Opinion in Neurobiology*, 37, 66-74.
- Gabbott, P. L., Warner, T. A., Jays, P. R., Salway, P., & Busby, S. J. (2005). Prefrontal cortex in the rat: projections to subcortical autonomic, motor, and limbic centers. *Journal of Comparative Neurology*, 492(2), 145-177.
- Genud-Gabai, R., Klavir, O., & Paz, R. (2013). Safety signals in the primate amygdala. *Journal of Neuroscience*, 33(46), 17986-17994.
- Ghosh, S., & Chattarji, S. (2015). Neuronal encoding of the switch from specific to generalized fear. *Nature Neuroscience*, 18(1), 112-120.
- Gini, C. (1921). Measurement of inequality of incomes. *The Economic Journal*, 31, 124-126.
- Giustino, T. F., & Maren, S. (2015). The Role of the Medial Prefrontal Cortex in the Conditioning and Extinction of Fear. *Frontiers in Behavioral Neuroscience*, 9, 298.
- Grewe, B. F., Grundemann, J., Kitch, L. J., Lecoq, J. A., Parker, J. G., Marshall, J. D., . . . Schnitzer, M. J. (2017). Neural ensemble dynamics underlying a long-term associative memory. *Nature*, 543(7647), 670-675.

- Grundemann, J., Bitterman, Y., Lu, T., Krabbe, S., Grewe, B. F., Schnitzer, M. J., & Luthi, A. (2019). Amygdala ensembles encode behavioral states. *Science*, 364(6437).
- Halladay, L. R., & Blair, H. T. (2015). Distinct ensembles of medial prefrontal cortex neurons are activated by threatening stimuli that elicit excitation vs. inhibition of movement. *Journal of Neurophysiology*, 114(2), 793-807.
- Han, J. H., Kushner, S. A., Yiu, A. P., Cole, C. J., Matynia, A., Brown, R. A., . . . Josselyn, S. A. (2007). Neuronal competition and selection during memory formation. *Science*, 316(5823), 457-460.
- Haubensak, W., Kunwar, P. S., Cai, H., Ciocchi, S., Wall, N. R., Ponnusamy, R., . . . Anderson, D. J. (2010). Genetic dissection of an amygdala microcircuit that gates conditioned fear. *Nature*, 468(7321), 270-276.
- Hazan, L., Zugaro, M., & Buzsaki, G. (2006). Klusters, NeuroScope, NDManager: a free software suite for neurophysiological data processing and visualization. *Journal of Neuroscience Methods*, 155(2), 207-216.
- Headley, D. B., Kanta, V., Kyriazi, P., & Pare, D. (2019). Embracing Complexity in Defensive Networks. *Neuron*, 103(2), 189-201.
- Heidbreder, C. A., & Groenewegen, H. J. (2003). The medial prefrontal cortex in the rat: evidence for a dorso-ventral distinction based upon functional and anatomical characteristics. *Neuroscience and Biobehavioral Reviews*, 27(6), 555-579.
- Hintiryan, H., Bowman, I., Johnson, D. L., Korobkova, L., Zhu, M., Khanjani, N., . . . Dong, H.-W. (2019). Connectivity characterization of the mouse basolateral amygdalar complex. *bioRxiv*, 807743.
- Hoover, W. B., & Vertes, R. P. (2007). Anatomical analysis of afferent projections to the medial prefrontal cortex in the rat. *Brain Structure & Function*, 212(2), 149-179.
- Hyman, J. M., Ma, L., Balaguer-Ballester, E., Durstewitz, D., & Seamans, J. K. (2012). Contextual encoding by ensembles of medial prefrontal cortex neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 109(13), 5086-5091.
- Ishikawa, A., Ambroggi, F., Nicola, S. M., & Fields, H. L. (2008). Contributions of the amygdala and medial prefrontal cortex to incentive cue responding. *Neuroscience*, 155(3), 573-584.

- Kalin, N. H., & Shelton, S. E. (1989). Defensive behaviors in infant rhesus monkeys: environmental cues and neurochemical regulation. *Science*, 243(4899), 1718-1721.
- Keene, C. S., Bladon, J., McKenzie, S., Liu, C. D., O'Keefe, J., & Eichenbaum, H. (2016). Complementary Functional Organization of Neuronal Activity Patterns in the Perirhinal, Lateral Entorhinal, and Medial Entorhinal Cortices. *Journal of Neuroscience*, 36(13), 3660-3675.
- Kim, J., Pignatelli, M., Xu, S., Itohara, S., & Tonegawa, S. (2016). Antagonistic negative and positive neurons of the basolateral amygdala. *Nature Neuroscience*, 19(12), 1636-1646.
- Klavir, O., Genud-Gabai, R., & Paz, R. (2013). Functional connectivity between amygdala and cingulate cortex for adaptive aversive learning. *Neuron*, 80(5), 1290-1300.
- Kluver, H., & Bucy, P. C. (1997). Preliminary analysis of functions of the temporal lobes in monkeys. 1939. *Journal of Neuropsychiatry and Clinical Neurosciences*, 9(4), 606-620.
- Krettek, J. E., & Price, J. L. (1977). Projections from the amygdaloid complex to the cerebral cortex and thalamus in the rat and cat. *Journal of Comparative Neurology*, 172(4), 687-722.
- Krettek, J. E., & Price, J. L. (1978). Amygdaloid projections to subcortical structures within the basal forebrain and brainstem in the rat and cat. *Journal of Comparative Neurology*, 178(2), 225-254.
- Kyriazi, P., Headley, D. B., & Pare, D. (2018). Multi-dimensional Coding by Basolateral Amygdala Neurons. *Neuron*, 99(6), 1315-1328.e1315.
- Laurent, V., & Westbrook, R. F. (2009). Inactivation of the infralimbic but not the prelimbic cortex impairs consolidation and retrieval of fear extinction. *Learning and Memory*, 16(9), 520-529.
- Lazaro-Munoz, G., LeDoux, J. E., & Cain, C. K. (2010). Sidman instrumental avoidance initially depends on lateral and basal amygdala and is constrained by central amygdala-mediated Pavlovian processes. *Biological Psychiatry*, 67(12), 1120-1127.
- LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, 23, 155-184.

- LeDoux, J. E., Farb, C., & Ruggiero, D. A. (1990). Topographic organization of neurons in the acoustic thalamus that project to the amygdala. *Journal of Neuroscience*, 10(4), 1043-1054.
- Lee, S. C., Amir, A., Haufler, D., & Pare, D. (2017). Differential Recruitment of Competing Valence-Related Amygdala Networks during Anxiety. *Neuron*, 96(1), 81-88.e85.
- Lee, S. C., Amir, A., Headley, D. B., Haufler, D., & Pare, D. (2016). Basolateral amygdala nucleus responses to appetitive conditioned stimuli correlate with variations in conditioned behaviour. *Nat Commun*, 7, 12275.
- Lindsay, G. W., Rigotti, M., Warden, M. R., Miller, E. K., & Fusi, S. (2017). Hebbian Learning in a Random Network Captures Selectivity Properties of the Prefrontal Cortex. *Journal of Neuroscience*, 37(45), 11021-11036.
- Little, J. P., & Carter, A. G. (2013). Synaptic mechanisms underlying strong reciprocal connectivity between the medial prefrontal cortex and basolateral amygdala. *Journal of Neuroscience*, 33(39), 15333-15342.
- Livneh, U., Resnik, J., Shohat, Y., & Paz, R. (2012). Self-monitoring of social facial expressions in the primate amygdala and cingulate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 109(46), 18956-18961.
- Ma, L., Hyman, J. M., Durstewitz, D., Phillips, A. G., & Seamans, J. K. (2016). A Quantitative Analysis of Context-Dependent Remapping of Medial Frontal Cortex Neurons and Ensembles. *Journal of Neuroscience*, 36(31), 8258-8272.
- Ma, L., Hyman, J. M., Lindsay, A. J., Phillips, A. G., & Seamans, J. K. (2014). Differences in the emergent coding properties of cortical and striatal ensembles. *Nature Neuroscience*, 17(8), 1100-1106.
- Maren, S., & Quirk, G. J. (2004). Neuronal signalling of fear memory. *Nature Reviews: Neuroscience*, 5(11), 844-852.
- McDonald, A. J. (1991). Organization of amygdaloid projections to the prefrontal cortex and associated striatum in the rat. *Neuroscience*, 44(1), 1-14.
- McDonald, A. J. (1992). Cell types and intrinsic connections of the amygdala. . In *The amygdala: Neurobiological aspects of emotion, memory, and mental dysfunction*, J.P. Aggleton, ed. (New York: Wiley-Liss). 67-96.

- McDonald, A. J. (1998). Cortical pathways to the mammalian amygdala. *Progress in Neurobiology*, 55(3), 257-332.
- McDonald, A. J., Mascagni, F., & Guo, L. (1996). Projections of the medial and lateral prefrontal cortices to the amygdala: a Phaseolus vulgaris leucoagglutinin study in the rat. *Neuroscience*, 71(1), 55-75.
- McGarry, L. M., & Carter, A. G. (2017). Prefrontal Cortex Drives Distinct Projection Neurons in the Basolateral Amygdala. *Cell Reports*, 21(6), 1426-1433.
- McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Riviere, P. D., & Eichenbaum, H. (2014). Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron*, 83(1), 202-215.
- McKernan, M. G., & Shinnick-Gallagher, P. (1997). Fear conditioning induces a lasting potentiation of synaptic currents in vitro. *Nature*, 390(6660), 607-611.
- Meyers, E. M., Freedman, D. J., Kreiman, G., Miller, E. K., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology*, 100(3), 1407-1419.
- Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews: Neuroscience*, 1(1), 59-65.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167-202.
- Mobbs, D., Headley, D. B., Ding, W., & Dayan, P. (2020). Space, Time, and Fear: Survival Computations along Defensive Circuits. *Trends in Cognitive Sciences*, 24(3), 228-241.
- Mobbs, D., Yu, R., Rowe, J. B., Eich, H., FeldmanHall, O., & Dalgleish, T. (2010). Neural activity associated with monitoring the oscillating threat value of a tarantula. *Proceedings of the National Academy of Sciences of the United States of America*, 107(47), 20582-20586.
- Moorman, D. E., & Aston-Jones, G. (2015). Prefrontal neurons encode context-based response execution and inhibition in reward seeking and extinction. *Proceedings of the National Academy of Sciences of the United States of America*, 112(30), 9472-9477.

- Morris, A. P., Bremmer, F., & Krekelberg, B. (2016). The Dorsal Visual System Predicts Future and Remembers Past Eye Position. *Frontiers in Systems Neuroscience*, 10, 9.
- Moscarello, J. M., & LeDoux, J. E. (2013). Active avoidance learning requires prefrontal suppression of amygdala-mediated defensive reactions. *Journal of Neuroscience*, 33(9), 3815-3823.
- Muller, J. F., Mascagni, F., & McDonald, A. J. (2006). Pyramidal cells of the rat basolateral amygdala: synaptology and innervation by parvalbumin-immunoreactive interneurons. *Journal of Comparative Neurology*, 494(4), 635-650.
- Munuera, J., Rigotti, M., & Salzman, C. D. (2018). Shared neural coding for social hierarchy and reward value in primate amygdala. *Nature Neuroscience*, 21(3), 415-423.
- Namburi, P., Beyeler, A., Yorozu, S., Calhoon, G. G., Halbert, S. A., Wichmann, R., . . . Tye, K. M. (2015). A circuit mechanism for differentiating positive and negative associations. *Nature*, 520(7549), 675-678.
- Ongur, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex*, 10(3), 206-219.
- Otis, J. M., Namboodiri, V. M., Matan, A. M., Voets, E. S., Mohorn, E. P., Kosyk, O., . . . Stuber, G. D. (2017). Prefrontal cortex output circuits guide reward seeking through divergent cue encoding. *Nature*, 543(7643), 103-107.
- Pachitariu, M., Steinmetz, N. A., Kadir, S., Carandini, M., & Harris, K. D. (2016). Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv*.
- Pape, H. C., & Pare, D. (2010). Plastic synaptic networks of the amygdala for the acquisition, expression, and extinction of conditioned fear. *Physiological Reviews*, 90(2), 419-463.
- Pare, D., & Quirk, G. J. (2017). WHEN SCIENTIFIC PARADIGMS LEAD TO TUNNEL VISION: LESSONS FROM THE STUDY OF FEAR. *NPJ Sci Learn*, 2.
- Pare, D., & Smith, Y. (1993). Distribution of GABA immunoreactivity in the amygdaloid complex of the cat. *Neuroscience*, 57(4), 1061-1076.

- Park, I. M., Meister, M. L., Huk, A. C., & Pillow, J. W. (2014). Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nature Neuroscience*, 17(10), 1395-1403.
- Paton, J. J., Belova, M. A., Morrison, S. E., & Salzman, C. D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature*, 439(7078), 865-870.
- Pellman, B. A., & Kim, J. J. (2016). What Can Ethobehavioral Studies Tell Us about the Brain's Fear System? *Trends in Neurosciences*, 39(6), 420-431.
- Perusini, J. N., & Fanselow, M. S. (2015). Neurobehavioral perspectives on the distinction between fear and anxiety. *Learning and Memory*, 22(9), 417-425.
- Peters, J., Kalivas, P. W., & Quirk, G. J. (2009). Extinction circuits for fear and addiction overlap in prefrontal cortex. *Learning and Memory*, 16(5), 279-288.
- Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. *Neuron*, 48(2), 175-187.
- Pitkanen, A. (2000). Connectivity of the rat amygdaloid complex. In *The Amygdala: A Functional Analysis*, John P. Aggleton, ed. (New York: Oxford University Press), 31-116.
- Pitkänen, A. (2000). Connectivity of the rat amygdaloid complex. In: *Aggleton, J.P., Ed., The Amygdala*, Oxford University Press, New York, 31-116.
- Pitkanen, A., Savander, V., & LeDoux, J. E. (1997). Organization of intra-amygdaloid circuitries in the rat: an emerging framework for understanding functions of the amygdala. *Trends in Neurosciences*, 20(11), 517-523.
- Poirazi, P., Brannon, T., & Mel, B. W. (2003). Pyramidal Neuron as Two-Layer Neural Network. *Neuron*, 37, 989-999.
- Pryluk, R., Kfir, Y., Gelbard-Sagiv, H., Fried, I., & Paz, R. (2019). A Tradeoff in the Neural Code across Regions and Species. *Cell*, 176(3), 597-609.e518.
- Quirk, G. J., Repa, C., & LeDoux, J. E. (1995). Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: parallel recordings in the freely behaving rat. *Neuron*, 15(5), 1029-1039.

- Ray, J. P., & Price, J. L. (1992). The organization of the thalamocortical connections of the mediodorsal thalamic nucleus in the rat, related to the ventral forebrain-prefrontal cortex topography. *Journal of Comparative Neurology*, 323(2), 167-197.
- Repa, J. C., Muller, J., Apergis, J., Desrochers, T. M., Zhou, Y., & LeDoux, J. E. (2001). Two different lateral amygdala cell populations contribute to the initiation and storage of memory. *Nat Neurosci*, 4(7), 724-731.
- Rigotti, M., Barak, O., Warden, M. R., Wang, X. J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451), 585-590.
- Roelofs, K. (2017). Freeze for action: neurobiological mechanisms in animal and human freezing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 372(1718).
- Romanski, L. M., Clugnet, M. C., Bordi, F., & LeDoux, J. E. (1993). Somatosensory and auditory convergence in the lateral nucleus of the amygdala. *Behavioral Neuroscience*, 107(3), 444-450.
- Rozeske, R. R., Jercog, D., Karalis, N., Chaudun, F., Khoder, S., Girard, D., . . . Herry, C. (2018). Prefrontal-Periaqueductal Gray-Projecting Neurons Mediate Context Fear Discrimination. *Neuron*, 97(4), 898-910.e896.
- Rumpel, S., LeDoux, J., Zador, A., & Malinow, R. (2005). Postsynaptic receptor trafficking underlying a form of associative learning. *Science*, 308(5718), 83-88.
- Saez, A., Rigotti, M., Ostojic, S., Fusi, S., & Salzman, C. D. (2015). Abstract Context Representations in Primate Amygdala and Prefrontal Cortex. *Neuron*, 87(4), 869-881.
- Saez, R. A., Saez, A., Paton, J. J., Lau, B., & Salzman, C. D. (2017). Distinct Roles for the Amygdala and Orbitofrontal Cortex in Representing the Relative Amount of Expected Reward. *Neuron*, 95(1), 70-77.e73.
- Sah, P., Faber, E. S., Lopez De Armentia, M., & Power, J. (2003). The amygdaloid complex: anatomy and physiology. *Physiological Reviews*, 83(3), 803-834.
- Sangha, S., Chadick, J. Z., & Janak, P. H. (2013). Safety encoding in the basal amygdala. *Journal of Neuroscience*, 33(9), 3744-3751.

- Sangha, S., Diehl, M. M., Bergstrom, H. C., & Drew, M. R. (2020). Know safety, no fear. *Neuroscience and Biobehavioral Reviews*, 108, 218-230.
- Savander, V., Go, C. G., LeDoux, J. E., & Pitkanen, A. (1995). Intrinsic connections of the rat amygdaloid complex: projections originating in the basal nucleus. *Journal of Comparative Neurology*, 361(2), 345-368.
- Sesack, S. R., Deutch, A. Y., Roth, R. H., & Bunney, B. S. (1989). Topographical organization of the efferent projections of the medial prefrontal cortex in the rat: an anterograde tract-tracing study with Phaseolus vulgaris leucoagglutinin. *Journal of Comparative Neurology*, 290(2), 213-242.
- Shabel, S. J., & Janak, P. H. (2009). Substantial similarity in amygdala neuronal activity during conditioned appetitive and aversive emotional arousal. *Proceedings of the National Academy of Sciences of the United States of America*, 106(35), 15031-15036.
- Sierra-Mercado, D., Jr., Corcoran, K. A., Lebron-Milad, K., & Quirk, G. J. (2006). Inactivation of the ventromedial prefrontal cortex reduces expression of conditioned fear and impairs subsequent recall of extinction. *European Journal of Neuroscience*, 24(6), 1751-1758.
- Sierra-Mercado, D., Padilla-Coreano, N., & Quirk, G. J. (2011). Dissociable roles of prelimbic and infralimbic cortices, ventral hippocampus, and basolateral amygdala in the expression and extinction of conditioned fear. *Neuropsychopharmacology*, 36(2), 529-538.
- Sotres-Bayon, F., & Quirk, G. J. (2010). Prefrontal control of fear: more than just extinction. *Current Opinion in Neurobiology*, 20(2), 231-235.
- Swanson, L. W. (1981). A direct projection from Ammon's horn to prefrontal cortex in the rat. *Brain Research*, 217(1), 150-154.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J R Statist Soc B*, 58, 267-288.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1), 267-288.
- Tye, K. M., Stuber, G. D., de Ridder, B., Bonci, A., & Janak, P. H. (2008). Rapid strengthening of thalamo-amygdala synapses mediates cue-reward learning. *Nature*, 453(7199), 1253-1257.

- Vertes, R. P. (2004). Differential projections of the infralimbic and prelimbic cortex in the rat. *Synapse*, 51(1), 32-58.
- Vertes, R. P., & Hoover, W. B. (2008). Projections of the paraventricular and paratenial nuclei of the dorsal midline thalamus in the rat. *Journal of Comparative Neurology*, 508(2), 212-237.
- Vetere, G., Kenney, J. W., Tran, L. M., Xia, F., Steadman, P. E., Parkinson, J., . . . Frankland, P. W. (2017). Chemogenetic Interrogation of a Brain-wide Fear Memory Network in Mice. *Neuron*, 94(2), 363-374.e364.
- Wall, N. R., De La Parra, M., Callaway, E. M., & Kreitzer, A. C. (2013). Differential innervation of direct- and indirect-pathway striatal projection neurons. *Neuron*, 79(2), 347-360.
- Weiskrantz, L. (1956). Behavioral changes associated with ablation of the amygdaloid complex in monkeys. *Journal of Comparative and Physiological Psychology*, 49(4), 381-391.
- Yiu, A. P., Mercaldo, V., Yan, C., Richards, B., Rashid, A. J., Hsiang, H. L., . . . Josselyn, S. A. (2014). Neurons are recruited to a memory trace based on relative neuronal excitability immediately before training. *Neuron*, 83(3), 722-735.
- Youngentob, S. L., Johnson, B. A., Leon, M., Sheehe, P. R., & Kent, P. F. (2006). Predicting odorant quality perceptions from multidimensional scaling of olfactory bulb glomerular activity patterns. *Behavioral Neuroscience*, 120(6), 1337-1345.
- Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J R Statist Soc B*, 68, 49-67.
- Yuan, M., & Lin, Y. (2006). Model Selection and Estimation in Regression With Grouped Variables. *Journal of the Royal Statistical Society Series B*, 68, 49-67.
- Zhang, X., & Li, B. (2018). Population coding of valence in the basolateral amygdala. *Nat Commun*, 9(1), 5195.