Midbrain Dopamine Neurons During Appetitive and Aversive States

Anna-Lena B. Schlenner

A Thesis

in

The Department

of

Psychology

Presented in partial fulfillment of the requirements for the

Degree of Master of Arts (Psychology)

at Concordia University

Montreal, Quebec, Canada

November 2021

Concordia University

School of Graduate Studies

This is to certify that the thesis prepared

By: Anna-Lena B. Schlenner

Entitled: Midbrain dopamine neurons during appetitive and aversive states

and submitted in partial fulfillment of the requirements for the degree of

Master of Arts (Psychology)

Complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____ Chair
Dr. Kristen Dunfield

_____ Examiner
Dr. Andrew Chapman

_____ Examiner
Dr. Matthew Gardner

_____ Supervisor
Dr. Mihaela Iordanova

Approved by _____
Dr. Kristen Dunfield

_____ 2021 _____
Dr. Pascale Sicotte (Dean of Faculty)

Abstract

Midbrain Dopamine Neurons During Appetitive and Aversive States

Anna-Lena B. Schlenner

A key role ascribed to midbrain dopamine (DA) neurons rests with learning about rewarding events by reflecting reward prediction errors (RPE). Research has shown that during reward learning a positive prediction error (e.g. surprising reward), leads to phasic excitation, while a negative prediction error (e.g. omission of an expected reward) leads to phasic inhibition in DA neurons. It remains unclear, however, how DA regulates learning about aversive events. Using behavioral electrophysiology we recorded from DA neurons in the ventral tegmental area (VTA) during a Pavlovian task in which auditory cues were trained as predictors of either an appetitive sucrose reward or aversive footshock. Our analyses confirmed a role for VTA DA neurons in tracking reward prediction error (RPE), that is, elevation in firing rate (FR) to the reward predictor and depression in FR at time of reward omission in a correlated fashion. Further, our goal was to determine whether DA firing would represent reward and aversion in line with a valence-based prediction error signal. We found that cue related phasic DA activity to both reward and aversion predicting cues contained both information about stimulus identity, as well as valence. Additionally, outcome omission was represented as state of opposite valence. These results support the hypothesis that midbrain DA neurons support learning by signaling valence-based prediction errors.

Acknowledgements

I would like to thank my supervisor Dr. Mihaela Iordanova for her support and guidance through the project and Dr. Matt Gardner for helpful discussions regarding data analysis.

In addition, I thank the Natural Sciences and Engineering Research Council of Canada, as well as the Canadian Research Chairs for funding my work. Thank you to Concordia University for support through the International Tuition Remission Award and the Faculty of Arts and Sciences Award for Academic Excellence.

To my defense committee Dr. Mihaela Iordanova, Dr. Andrew Chapman, and Dr. Matt Gardner, thank you for your time and effort in helping me realize my dream.

Thank you to all my lab and class mates for their laughs, hugs, and socially distanced game nights.

Mom, Dad, Tara, and Jonathan, thank you so much for always being there for me, for keeping the family bonds strong even from afar, and teaching me scientific curiosity from a young age.

Ken, thank you for being my light and my rock through this time, for sharing your love with me and for your incredible ability to always put a smile on my face.

Contribution of Authors

Dr. Mihaela Iordanova designed the experimental design and obtained funding. Anna-Lena Schlenner conducted the experiment and all associated tasks, analyzed the resulting data, and wrote the manuscript under the guidance of Mihaela Iordanova.

Table of Contents

List of Figures

Introduction

Associative Learning Through Reward Prediction Errors.

In traditional Pavlovian conditioning experiments, a neutral cue (conditioned stimulus [CS]) such as a light or tone is paired with an outcome of biological significance (unconditioned stimulus [US]) of biological relevance. This could either be a rewarding event like a food pellet or an aversive event like a foot shock. Such pairings establish the cue as a predictor for the outcome and thus elicits conditioned responding. This could be observed as an approach behavior to where food is delivered, or depression of ongoing behavior in anticipation of the aversive event. While classical conditioning emphasizes the contiguous presentation of the cue and outcome, the field of associative learning has long known that the establishment of an association between cue and outcome critically depends on prediction error. In the absence of prediction error, even temporal contiguity does not ensure association formation.

Prediction error (PE) is defined by the discrepancy between received and expected outcomes (Rescorla & Wagner, 1972). Early in learning PE is maximal, leading to large changes in associative strength. As the associative strength increases, that is, late in learning, the discrepancy between the expectation and the outcome is reduced, leading to smaller or no changes in learning. This means the increments in trial-by-trial learning become smaller as training progresses. Additionally, the sign (positive or negative) of the error affects the directionality of the change in association. An outcome exceeding expectations results in a positive PE, leading to an increase in associative strength between cue and outcome. On the other hand, if an outcome is worse than expected, this causes a negative PE, resulting in a decrease in associative strength.

The blocking design best exemplifies the teaching role of PE in associative learning (Kamin, 1968, 1969). The blocking paradigm consists of two phases. During the first phase, a cue (A) is conditioned to predict an outcome (e.g., sugar pellets, footshock). During the second phase, A is presented simultaneously with a novel cue (X) and reinforced with the same outcome as during phase 1. To determine how much associative strength had accrued to the blocked cue X, a control group is trained identically to the blocking groups with the exception of no phase 1 pretraining. In a within subjects design this means that the control compound consists of two novel stimuli, B and Y, reinforced with the same outcome as that used for AX. On test, Y elicits a stronger conditioned response compared to X, suggesting that learning about the association between the former cue and the outcome is greater compared to learning about the latter. It is said that the pretrained cue A blocked learning about the X→outcome association because it already predicted outcome delivery. In other works, there was little or no error in prediction in the blocking (AX) group, hindering the establishment of an association between X and the outcome. On the other hand, neither B nor Y had previously been associated with reward, therefore prediction error was maximal on BY conditioning trials, resulting in learning the Y→outcome (and B→outcome) association. Learning about the blocked cue X can be restored, however, by increasing the prediction error by delivering an unexpected (e.g. larger) outcome during AX training.

On the other hand, presenting a previously reinforced cue without reinforcement, generates a negative prediction error at the time of the outcome (Bouton & Bolles, 1979). Over multiple non-reinforced presentations of the cue, one can observe a continuous decline of the conditioned response (Rescorla & Wagner, 1972). This is called an extinction effect. A negative PE therefore results in a decline in associative strength, leading to reduction of the conditioned response.

These procedures provide insight into how alterations in environmental contingencies generate PEs and thereby influence learning and behavior. How the brain computes environmental input and regulates appropriate responses, however, cannot be understood from these studies. It is therefore crucial to investigate the neural mechanisms underlying the changes in behavior following errors in prediction.

Phasic firing of VTA DA neurons tracks reward PE (RPE).

Electrical stimulation of dopamine (DA) neurons in the ventral tegmental area (VTA) and substantia nigra pars compacta (SNc) has been found to reinforce behavior (Olds & Milner, 1954). Additionally, these neurons have been linked to tracking PEs in reward studies. Across species, VTA DA neurons show a bidirectional activity pattern that matches the sign of the PE signal (Eshel et al., 2016; Mirenowicz & Schultz, 1994; Pan et al., 2005). Electrophysiological studies have reported a brief phasic increase in firing rate at the time of an unexpected reward in accordance with a positive PE (Mirenowicz & Schultz, 1994; Waelti et al., 2001). As a result of pairings between the rewarding outcome and antecedent cue, the latter become reliable predictors of outcome delivery. In turn, the VTA DA neuron phasic activity shifts from the now well-predicted outcome to the cue that signals its delivery (Mirenowicz & Schultz, 1994; Waelti et al., 2001). When a negative PE is introduced by omitting an expected reward, the VTA DA neurons briefly cease their firing altogether (Schultz et al., 1997). While the magnitude of the neuronal responses change with in accordance with reward valence (e.g. probability (Fiorillo et al., 2003) and size (Bayer et al., 2007)), the underlying bidirectional activity patterns remain the same.

Waelti et al. (2001) demonstrated in their study, that the activity of individual dopamine neurons corresponds to the presence of prediction errors. They presented monkeys with two stimuli (A and B) during phase 1 of the blocking paradigm. However only A was

reinforced, leading to a conditioned response to A, but not B. During the blocking phase, A was presented together with X, and B with Y. Both compounds were reinforced equally. The researchers only observed an increase in dopamine firing at the time points that represented a PE. According to formal learning theories (Rescorla & Wagner, 1972; Sutton, 1988), these occur at the time of an unexpected outcome, like the reward following the presentation of BY. In addition, the PE backpropagates to the earliest reliable predictor, which was evidence in phasic firing of VTA DA neurons to the AX compound but not to the reward that followed this compound. On test, the researchers observed a phasic increase in dopamine activity at the onset of stimulus Y, but not to X. These patterns reflect the behavioral differences between blocked (X) and unblocked stimuli (Y). In addition, this non-reinforced test resulted in phasic VTA DA neuron firing inhibition at time of reward delivery following Y but not X, indicating that a reward was expected following Y but not X. The authors also conducted a reinforced test, which also revealed higher level of VTA DA phasic activity to reward delivery after the blocked stimulus X compared to Y. This suggests that receiving a reward was surprising following X, but not following Y. Overall, VTA DA phasic activity tracked positive and negative reward PEs in a bidirectional manner and this activity predicted associative learning.

Causal evidence for the role of VTA DA neuron activity in PE.

PEs are critical to learning. To determine whether the RPE profile of VTA DA neuron firing has a causal role in associative learning, Steinberg et al. (2013) used optogenetics to artificially boost the VTA DA signal at time of the expected reward in a blocking design. As Waelti et. al. (2001) found, no PE occurs at the time of the US following presentation of AX (A - conditioned, X - novel) during a regular blocking experiment. In the Steinberg et. al. (2013) study, they hypothesized that if the dopamine transient signals a PE, activating

dopamine neurons during the US would artificially introduce a PE, preventing blocking of X. Indeed, animals responded more strongly to X if stimulation had occurred during its paired US. This suggests that stimulation of dopamine neurons mimics a positive PE.

To fully link positive as well as negative PEs to the bidirectional dopamine activity, it needs to be causally shown that inhibition of dopamine activity results in the same changes in activity as a negative PEs. To test this, Chang et al. (2016) prevented the dopamine transient at the time of an unexpected additional reward during the blocking phase. They hypothesized that if inhibition of dopamine signals a negative PE, the blocking effect should remain intact, despite the additional reward. Their data showed comparable responding to the stimulus associated with the manipulated reward and a blocked control stimulus. This confirmed their hypothesis, suggesting the bidirectional dopamine activity does indeed represent bidirectional PEs.

Even though the phasic increase in dopamine activity at the onset of a conditioned stimulus has been observed across the board by studies investigating midbrain dopamine activity, the actual function of that activity pattern has been debated. Two hypotheses could account for the cue evoked phasic DA activity. The temporal difference learning theory (Sutton, 1988) states that the current state of the animal (t) is compared with the previous state ($t-1$). Therefore, the onset of a cue would generate a prediction error. The second hypothesis suggests that phasic DA activity at cue onset signals the prediction of a future reward. To distinguish between these two options, Maes et. al. (2020) conducted a blocking experiment. They optogenetically shunted dopamine activity during cue onset of the blocking phase. The authors hypothesized that preventing the increase in dopamine activity at the onset of a conditioned stimulus during the compound phase will leave blocking intact if the phasic activity indeed signals a cue-evoked prediction error. If on the other hand the phasic DA activity signaled not a prediction error, but was simply predicting a future reward,

blocking would be disrupted, leading to learning about the novel stimulus (X). They indeed found that shunting dopamine activity in the VTA at the onset of a conditioned stimulus leaves the blocking effect intact. These data therefore provide causal evidence for the hypothesis that the dopamine transient at the onset of a conditioned stimulus serves as a temporal difference prediction error.

Theory of opposing appetitive and aversive systems

Temporal difference learning theory is not restricted to reward learning. In fact, it can be applied to learning about any predicted outcomes, as long as predictor and outcome are temporally related. This means that both appetitive and aversive outcomes should be able to be predicted by temporal difference PEs. However, it would be of importance for a biological system to be able to distinguish between aversive and appetitive outcomes. A theory was brought forward by Dickinson and Dearing (1979) on how this can be achieved. They proposed appetitive and aversive systems that are integrated with each other in an inhibitory manner. They suggested that the appetitive system would get activated by an attractive excitor like an unexpected reward or a stimulus predicting a reward. The aversive system on the other hand would get activated by an aversive excitor like an unexpected aversive event or a stimulus predicting such event. At the same time, the omission of an anticipated event would be expected to activate the opposing motivational system: Omission of a reward (frustration) would induce an aversive state, while omission of a shock (relief) would induce an appetitive state. Behavioral experiments, both Pavlovian and instrumental, support this opponent-state model. One example that has been successfully demonstrated is counterconditioning.

During counterconditioning a stimulus is first conditioned to predict an appetitive outcome. Hereby, the conditioned stimulus becomes a conditioned excitor for the appetitive

system, eliciting an appetitive conditioned response such as salivating and swallowing (Konorski & Szwejkowska, 1956; Scavio, 1974), or entries into the food port (Nasser & McNally, 2012). In the next stage, the same stimulus then precedes an aversive outcome such as a footshock. Comparing the speed of acquisition of a conditioned fear response of this counterconditioned stimulus to a naïve stimulus that has only been paired with shock or a stimulus which was previously presented non-reinforced shows a retardation effect. Animals are slower to acquire the fear response to a cue that has been trained as a good predictor for reward. The same is observed when the stimulus is first trained to predict an aversive outcome, conditioned excitor of the aversive motivational state, before coming to predict an appetitive outcome. This retardation in the acquisition of a conditioned response supports the idea that the two opposing states are integrated in an inhibitory manner. A reward predicting stimulus would activate the appetitive system, preventing activation of the opposite, aversive system, and thereby preventing a defensive response appropriate for an aversive outcome. Similarly, a conditioned inhibitor of one motivational state serves as facilitator for the acquisition of a conditioned response representing the state of opposite valence, as seen in superconditioning (Dickinson, 1977; Nasser & McNally, 2012; Rescorla, 1981).

Superconditioning occurs when a novel cue is presented in compound with a conditioned inhibitor (Rescorla, 1981) or a conditioned excitor of opposite valence (Dickinson, 1977), resulting in increased levels of the conditioned response. PE theory suggests that contrasting to a classical blocking design, learning to the novel cue is here extrapolated rather than blocked. When instead of receiving an anticipated reward (negative appetitive PE) an aversive outcome occurs (positive aversive PE), the two errors are added, leading to a larger PE. In their superconditioning experiment, Nasser and McNally (2012) trained rats to expect a food reward following a light cue. During the next stage of the experiment, the light was presented in compound with a novel tone cue and reinforced with

an aversive footshock. When the conditioned freezing response was evaluated to non-reinforced presentations of the tone cue alone, elevated levels of freezing were observed as compared to a control group that had received fear conditioning to only the tone cue.

While this behavioral evidence suggests an interaction between appetitive and aversive motivational states underlying learning, this does not discount the prediction error theory. Combining the two theories might be able to paint a completer and more accurate picture. Establishing a stimulus as conditioned excitor for the appetitive motivational system could lead to a positive, appetitive PE, while associating another stimulus with the aversive motivational system could result in a positive, aversive PE. Relating this to neural activity, neurons representing an appetitive motivational system would be expected to be phasically excited by a reward predicting stimulus, while neurons representing an aversive motivational system would be anticipated to be phasically excited by cues predicting aversive outcomes.

Dopamine Activity During Aversion.

Dopamine neurons have been studied extensively during reward learning, but less is known about their firing patterns during aversive learning. Investigating the dopaminergic response during aversive learning will provide us with necessary insight into whether the PE signal carried by VTA DA neurons is exclusive to reward or extends to the aversive domain. Midbrain DA neurons could only represent reward, but not aversion. Another option could be modulation to both reward and aversion in a valence independent manner. The third possibility would be a valence-based PE signal, with DA neurons representing reward and aversion with oppositely signed PEs.

DA neurons only signal reward. A study by Fiorillo (2013) supports this single-dimension hypothesis. The author found that midbrain DA neurons were excited to a surprising juice reward and inhibited to omission of a predicted reward, in line with a RPE signal. However,

DA neurons showed no difference in modulation to delivery of aversive air puffs and a bitter solution compared to the omission of these aversive outcomes. Similarly, a second study (Mirenowicz & Schultz, 1996) found the majority of midbrain DA neurons did not respond to cues predicting aversive outcomes, while exhibiting phasic excitation to cues predicting appetitive outcomes. Low levels of activation to aversion predicting cues by a minority of neurons was attributed to generalization to appetitive stimuli. It needs to be noted that both studies used air puffs as the aversive outcome. An air puff might be considered less aversive than a footshock used by other studies, and therefore not elicit modulation in firing rates signaling an event of biological significance.

DA neurons signal valence independent PEs. Guarraci & Kapp (1999) found the majority of recorded VTA dopamine neurons responded with a significantly higher firing rate at the onset of a stimulus that predicted a footshock compared to a non-reinforced stimulus. This suggests that DA neurons are modulated not only by reward predicting, but also by aversion predicting cues. A study by Joshua et al. (2008) compared activity patterns of dopamine neurons in the midbrain to unexpected aversive air puffs as well as omitted air puffs. They found the neurons seemed to be activated by both aversive and appetitive outcomes, even though the magnitude of excitation was larger for the food reward. At omission of both outcomes, dopamine neurons did not show a significant modulation. Similarly, Matsumoto and Hikosaka (2009b) reported that a large number of putative DA neurons recorded in the VTA and SNc responded similarly to reward and aversion. They found that the majority of recorded neurons were excited by both unexpected juice rewards and unexpected air puffs. In addition, DA neurons also exhibited phasic excitation at the onset of both reward and aversion predicting cues. Only the probability of receiving an outcome changed the magnitude in modulation. Additionally, studies have found an increase in DA release downstream from the VTA following both a water reward and an aversive foot shock (Sorg

& Kalivas, 1991; Young et al., 1993). The same was found to stimuli predicting appetitive and aversive outcomes. Using a microdialysis technique, extracellular levels of DA in the nucleus accumbens (NAcc) were found to be elevated to stimuli predicting a footshock (Young et al., 1993). These findings suggest that midbrain DA neurons signal PEs independent of valence.

DA neurons represent valence-based PEs. The possibility that midbrain DA neurons signal signed PEs corresponding to valence has found most support. A study by Salinas-Hernandez et. al. (2018) conditioned mice to expect an aversive footshock following a 10 second auditory cue. They found that omission of the expected footshock during fear extinction elicited increased activity of dopamine neurons in the VTA, in line with a positive prediction error. In line with a negative PE, Ungless et al., (2004) observed inhibition of dopamine firing in response to an aversive tail pinch. A causal study in rats (Luo et al., 2018) found evidence that both inhibition at time of an aversive event and excitation at time of omission of the aversive event are required for learning. When optogenetically inhibiting VTA DA neurons at the time of an aversive footshock during fear learning, the researchers found increased fear response compared to a control group. Optogenetically shunting dopamine activity at the time of omission of an anticipated footshock, just like optogenetic activation at time of reward omission (Steinberg et al., 2013), in return led to retardation in extinction learning. This indicates the positive prediction error elicited by shock omission to be necessary for updating learned contingencies.

Based on the contradicting findings of the field, as well as results from their own study, Mileykovskiy and Morales (2011) came to the conclusion that the phasic dopamine signal is more complicated than a simple excitation or inhibition at the cue onset. They found dopamine neurons fell into three different categories of activity patterns. Some only showed inhibition to an aversive stimulus. Others responded with a slight inhibition at the onset as

well as offset of the cue and return to baseline firing during the majority of the stimulus. A third group of neurons showed initial excitation, similar to the response to an appetitive stimulus. However, this brief excitation was directly followed by inhibition. They emphasize that all three types of responses always included inhibition at some time during the cue.

Based on these prior findings, we wanted to investigate whether the midbrain DA signal to predictors of appetitive and aversive events, as well as to the omission of these events are in line with a valence-based PE signal. We therefore hypothesized that VTA DA neurons would show phasic excitation above baseline activity to appetitive excitors such as cues predicting reward, as well as the omission of an aversive outcome (footshock). At the same time, we hypothesized to find phasic inhibition to aversive excitors (e.g. aversion predicting stimulus or omission of a rewarding outcome). To achieve this, we recorded VTA DA neurons during a task in which two counterbalanced auditory stimuli were conditioned to either predict an appetitive outcome (sucrose pellets) or an aversive outcome (footshock). Following appetitive and aversive conditioning, we assessed phasic activity of VTA DA neurons during non-reinforced presentations of the two cues, as well as the time of outcome omission. Our goal was to determine how DA in the midbrain represents appetitive and aversive events relative to one another, and whether the activity patterns correspond with a valence-based PE signal. We predicted that the phasic activity at cue onset, which has previously been indicated to carry the error signal (Maes et al., 2020), provides information about stimulus identity. Additionally, we hypothesized that the phasic activity also contains valence information, and that omission of an anticipated outcome would be represented as the state of opposite valence.

Materials and Methods

Subjects

Seven male Long Evans rats were obtained from Charles River Laboratory at 3.5 months of age, weighing 300-350g. Rats were double housed prior to surgery and single housed from the date of surgery and throughout the experiment in a standard clear cage (44.5cm x 25.8cm x 21.7cm; with beta chip bedding). The animal cages were located in a climate-controlled colony room and maintained on a 12h reverse light-dark cycle (lights off at 9am). Food and water were available ad libitum prior to surgery and during recovery. Thereafter, the rats were food restricted to 85% of their preoperative weight. All experimental procedures took place during the dark cycle and were in accordance with the approval granted by the Canadian Council on Animal Care and the Concordia University Animal Research Ethics Committee.

Surgical Procedure

All rats underwent surgery to implant a chronic electrode array consisting of 16 tetrodes (four 20µm tungsten wires spun together; California Fine Wire, Grover Beach, CA). Prior to surgery, the tetrode tips were electroplated with gold (Neuralynx, Bozeman, MT) at 1kHz to lower the impedance to 60kΩ using the nanoZ multi-electrode impedance tester (White Matter LLC., Seattle, WA) controlled by a computer (Windows OS) running the nanoZ software (nanoZ 1.4). The custom designed implant was configured so that 8 electrodes would reach the ventral tegmental area (VTA) at a 10 degree angle (see below for details), and the other 8 the nucleus accumbens (NAcc; data not included). During surgery, all animals were anesthetized with Isoflurane gas (1-2.5% in O_2), hair removed and placed in a stereotactic frame (David Kopf Instruments, Tujunga, CA). Before making the incision, animals were injected subcutaneously with 10mg/kg Rimadyl (Pfizer, Kirkland QC,

50mg/ml) as an analgesic. To target the VTA a craniotomy was created at 5.84 mm posterior and 2.1 mm lateral of bregma, and dura was completely removed. Upon placement of the electrode array, one drop of sterile 0.5% sodium alginate solution was combined with two drops of sterile 10% calcium chloride solution at the craniotomy to form a gelatinous seal. Electrodes were slowly lowered to 5.8mm DV. Eight stainless steel screws placed on the top and on the sides of the skull and dental cement secured the electrode array to the skull.

The rats were allowed to recover for 7 days during which they were handled, weighed, and received a 1ml oral dose of Cephalexin daily.

Electrode turning

In the days following the surgery, all tetrodes were slowly lowered 6.9mm DV. Upon the start of behavioral training, the electrodes were advanced 40-80μm every other day after the session. Tetrode locations were confirmed post-mortem.

Behavioral Apparatus

Behavioral procedures were conducted in four operant-training chambers (Med Associates, St. Albans, VT, USA). Each chamber measured 31.8 cm in height, 26.7 cm in depth and 25.4 cm in width. The modular left and right walls were made of aluminum, and the back wall, front door, and ceiling were made of clear Perspex. Their floors consisted of stainless-steel rods, 4 mm in diameter, spaced 15 mm apart, center to center, with a tray below the floor. A metal mesh and a metal door were attached to the outside of the behavioral chamber to form a Faraday cage in order to reduce electrical noise during recordings. Two auditory stimuli (see below) were delivered via two speakers mounted on opposite sides on the outside of the behavioral chamber. Food reward (see below) was delivered into a magazine. The magazine was equipped with an infrared sensor to detect head entries. Each chamber was enclosed in a ventilated cabinet, lined with sound-proofing foam. Illumination of the cabinet was achieved by an infrared light mounted at the back of the cabinet. Below

the light, a camera was mounted which was connected to a monitor located in another room of the laboratory where the behavior of each rat was observed by an experimenter. Delivery of stimuli and reinforcement was controlled by Med Associates software (Version 4.2) on a computer located outside the experimental room. Behavioral chamber, grid floor and tray were cleaned with water after each animal.

Stimuli

Two 10 sec long auditory stimuli, a 1kHz tone at 73dB and white noise at 75dB (measured inside the chamber) were generated by the Med Associates software and delivered via two speakers mounted on opposite sides on the outside of the behavioral chamber. Each auditory stimulus was reinforced by either two chocolate flavored sucrose pellets (Product# F07256, Bio-Serv, Flemington, NJ) delivered into the magazine, or a mild foot shock (0.5mA) delivered through the grid floor. Stimulus-reinforcement combinations were counterbalanced between animals.

Behavioral Procedures

One day prior to the start of behavioral training, animals received 20 chocolate-flavored sucrose pellets in their home cage in order to reduce neophobia.

Magazine training. All rats received magazine training in the behavioural chambers. Each session lasted 40min and consisted of one pellet delivery every 60s (40 pellets in total). No other stimuli were presented during this session. The animals had to consume 20 or more pellets to advance to Phase 1 – Appetitive Conditioning.

Phase 1 - Appetitive conditioning. All rats received Pavlovian conditioning for eight consecutive days between one of the auditory cues (tone or white noise; counterbalanced) and two sucrose pellets in the behavioural chambers. The session lasted 65min, and consisted of 16 conditioning trials, 3.5min apart on average (range: 180s-240s). Five minutes following placement in the chamber (acclimation), the first conditioning trial began, and five min

following the last conditioning trial the animals were removed from the chambers. Each conditioning trial consisted of a 10s auditory cue (A) followed by the delivery of 2 sucrose pellets at 9.5s and 10.5s after cue onset.

Phase 2 Appetitive and Aversive (fear) conditioning. On days 8-12, rats received appetitive (as described above) and aversive (fear) conditioning. The session lasted 75-80 min and consisted of 16 appetitive and 4 aversive conditioning trials on average 3.5 min apart (range: 180-240sec). Five minutes following placement in the chamber (acclimation), the first conditioning trial began, and five minutes following the last conditioning trial the animal were removed from the chambers. Appetitive and aversive trials were presented in a pseudo-randomized order, which assured the occurrence of 1 aversive trial for 5 appetitive trials and prevented the clustering of all aversive trials. Aversive trials consisted of a 10sec novel auditory cue B (tone of white noise; counterbalanced) followed by a 0.5 sec long, 0.5mA foot shock at 9.5 sec after cue onset.

Test. On Day 13 all rats received a Test session, which consisted of four reminder trials (three appetitive and one aversive, in the order AABA) followed by 16 non-reinforced test trials. During the non-reinforced test trials each cue was presented alone in the absence of its associated outcome 8 times in a pseudo-randomized order (ABBABAAB etc.). The acclimation and cool-down period, total number of trials, and the ITI were identical to Phase 2 above.

The 5-day period of Phase 2 and Test was repeated 1-4 times per rat. For each repetition, the stimulus-outcome pairings remained consistent within animal. Only neuronal activity recorded during Test was analyzed.

Single-Unit Recordings

Extracellular in-vivo recordings were conducted using a custom-built electrode array. The array contained a total of 16 tetrodes (four 20µm tungsten wires spun together; California Fine Wire, Grover Beach, CA). Two individually drivable bundles containing 8 tetrodes each were located to reach the VTA and NAcc. The neuronal signals and their timestamps were amplified and digitized at the headstage and recorded at 25kHz using an eCube recording system (White Matter LLC, Seattle, WA). Real-time neural activity was monitored by the experimenter on a computer outside the behavior room using Open Ephys software (Siegle et al., 2017).

Spike Sorting

The continuous recorded signal was sorted offline into single units using the automated clustering framework Kilsort (https://github.com/cortex-lab/KiloSort; Pachitariu et al., 2016) in Matlab2016b (MathWorks Inc.). All clusters identified by the Kilosort algorithm were manually identified as single units using the Phy GUI-template (Rossant et al., 2016) and the following criteria: Single units had to be well isolated from the noise floor (amplitude of detected spikes formed at least 90% of a normal distribution) and had to produce an autocorrelation in accordance with an inter-spike interval. Only clusters that met all mentioned criteria were used in subsequent analyses.

Post-mortem Histology

After completing the experiment, all rats received a lethal dose of sodium pentobarbital diluted 1:1 with 0.9% sodium chloride (1ml). Rats were then perfused using 200 ml of 0.9% sodium chloride solution and 150ml of 4% paraformaldehyde solution. Two hours after perfusion, the electrodes were retracted from their final recording location and the brains removed. Sequential, coronal sections of the VTA region were collected and stained

with cresyl violet. The final location of the electrode tips was determined under a microscope using the boundaries defined by the rat brain atlas (Paxinos & Watson, 2006)

Quantification

Behavioural data: Conditioned responding was measured by the number of head entries made into the magazine during the 9.5 sec of the cue preceding reward delivery. Percent time spent in the magazine was calculated for each trial for each phase of the experiment. Response ratio was calculated by comparing responding during the cue to baseline, which was defined as the average responding across all trials during the 60s period preceding cue onset (CS/(CS+ITI)).

Neural data: All neuronal data were analyzed using Matlab (MathWorks Inc.). Dopamine neurons were identified by analyzing the neuronal waveforms as done previously (Roesch et al., 2007; Takahashi et al., 2011, 2017). Neurons were clustered according to their amplitude ratio (amplitude of first negative and first positive segment) and half duration (time from trough to second peak). Neurons that fell within 2SD of the cluster with smaller amplitude ratio and longer half duration were considered dopaminergic. Neurons that fell within 2SD of both clusters were not considered. Firing rate of these neurons was normalized to a 10 sec baseline period (-20 to -10 sec before cue onset). Baseline firing was first averaged across trials for each 100ms bin and then a mean and SD was calculated across the 100 bins.

Behavioral analyses and statistical tests

Repeated measures ANOVAs were used to analyze changes in behavior across days during Phase 1, as well as across days and between stimuli for Phase 2. A paired t-test was performed to compare responding between the two stimuli during the Test session.

Neural analyses and statistical tests

Firing rate analysis: In order to normalize firing rates, the pre-cue period (-10 to 0 sec) was divided into 100ms bins, and each bin normalized to the baseline values [(pre_cue_bin - baseline_mean)/baseline_SD]. The same was done for the cue (0-10sec) and outcome period (10-15sec).

Cumulative sum: To identify periods of excitation and inhibition for individual dopamine neurons a cumulative sum analysis was used (*cumsum*). For this, the number of spikes that occurred during 100ms bins were added to the sum of spikes of all previous bins across the cue period (0-10sec). Therefore, the change in cumulative sum (slope) from one epoch to the next was steep when the FR during an epoch was high and flat when no spikes occurred during an epoch. In addition, the cumulative sum of the pre-cue period (-10 to 0 sec) was calculated and bootstrapped 200 times. A 99% confidence interval (CI) was calculated for each time step of the bootstrapped pre-cue period and the slope of the cue period compared to the change in lower and upper bounds of the CI. If the cue slope was shallower than the slope of the lower bound of the CI for at least three consecutive 100ms bins, the neuron was considered inhibited for this period, if the cue slope was steeper than the slope of the upper bound for at least three consecutive 100ms bins, the period was considered excited.

Cue decoder models: For each of 1,000 permutations, a random subset of 40 dopamine neurons was selected. For each of the one hundred 100ms epochs of the 10 sec cue the following was performed: A principal component analysis (PCA) was conducted on the FR during the epoch for the subset of randomly selected neurons. The mean-centered FR of the cue epoch was mapped on the coefficients of the top 8 principal components (PCs). An Elastic Net regularized logistic generalized linear model (GLM; *lassoglm*) and a linear discriminant analysis (LDA; *fitcdiscr*) were then used to decode stimulus type. Maximum-likelihood for the GLM was determined as follows:

$$\min_{\beta_0,\beta}\left(\frac{1}{N}Deviance(\beta_0,\beta)+\lambda P_\alpha(\beta)\right),$$

where

$$P_\alpha(\beta)=\frac{(1-\alpha)}{2}||\beta||_2^2+\alpha||\beta||_1=\sum_{j=1}^{p}\left(\frac{(1-\alpha)}{2}\beta_j^2+\alpha|\beta_j|\right)$$

with α = 0.5.

- N is the number of observations
- β_0 is the intercept
- β are the predictor coefficients
- λ is the nonnegative regularization parameter
- $P_\alpha(\beta)$ is the penalty term

Discriminant classification of the LDA was determined as follows:

$$\hat{y}=\arg\min_{y=1,\ldots,K}\sum_{k=1}^{K}\hat{P}(k|x)C(y|k)$$

where

- $\hat{y}$ is the predicted classification.
- K is the number of classes.
- $\hat{P}(k|x)$ is the posterior probability of class *k* of observation *x*.
- $C(y|k)$ is the cost of classifying an observation as *y* when its true class is *k*.

The GLM and LDA models were trained on 50% (8 out of 16 test trials) of trials using k-fold cross-validation. The algorithm trained on data from 7 trials and validated on the 8th trial.

This was repeated until each trial had been used for validation once (8-fold cross-validation), optimizing the β weights. Performance of the model was then tested on the held-out data and evaluated by calculating the percent of trials that were classified correctly. In addition, the models were compared to random validation models (see "Validation Models") to identify periods where model performance was better or worse than the random model.

Cross-temporal/outcome decoder: Permutations, selection of 40 neurons and PCA were conducted as for the cue decoder. An Elastic Net regularized classifier model (GLM; *lassoglm*) and a linear discriminant analysis (LDA; *fitcdiscr*) were used to predict stimulus type from neural activity during the outcome omission period. Maximum likelihood and discriminant classification were determined as in the cue decoder. The GLM and LDA outcome decoder models were trained on the FR of all 16 trials during the 500ms period of outcome omission and cross-validated 16 times. Performance of the model was tested on one hundred 100ms epochs of firing during the cue period and evaluated by comparing to a random validation model (see below).

Validation Models: To validate the performance of both cue and outcome decoders, the trained models were tested on 16 trials from epochs of the 10 seconds preceding cue onset. An average 95% CI was calculated across the 10 sec period. If the average cue or outcome decoder accuracy of a 100ms epoch fell above the upper bound of the CI, decoder performance was determined to be better than a random model. If the average accuracy fell below the lower bound of the CI, model performance was considered worse than a random model.

By-neuron GLM & identification of distinct decoding populations: To receive accuracy values for individual neurons, we trained an Elastic Net regularized GLM on outcome omission activity of each neuron and tested the model on activity from cue epochs of the same neuron. This was repeated for each neuron. Additionally, model performance was

validated by testing the model on randomized activity during the 10 sec pre-cue period. An average 95% CI was calculated for each neuron's validation model. Accuracy was then compared to the CI of the validation model. Neurons that performed better than the validation model (accuracy > upper CI bound) for at least one epoch between 300-800ms after cue onset were assigned to the population "Better". Neurons that had accuracy values below the CI for at least 100ms during the cue onset period were assigned to the population "Worse". To compare cue to outcome omission activity of the two populations, the inverse averaged FR of A and B trials was calculated. For this, the FR to B trials of each neuron was multiplied by -1 (inversed), added to the FR to A trials of that neuron and divided by 2 (averaged). This was done for both periods of cue and outcome omission and lead to one modulation value per neuron and time period.

Statistical tests: A two-sample t-test was conducted to identify the neurons with a significantly longer spike duration and smaller amplitude ratio (dopamine neurons; see "Neural Data"). A Spearman correlation was used to compare normalized firing rates (FRs), as well as the inversed average FRs of "Better" and "Worse" populations (see above) between cue and outcome omission periods. To compare the distribution of FRs between stimuli, as well as the duration of excitatory and inhibitory periods (see "Cumulative sum") a two-sample Kolmogorov-Smirnoff (KS) test was used. A Cox regression was utilized to compare differences in timing of events (time of maximum/minimum FR & onset of excitation/inhibition) between trial types. To compare the distribution of FRs to A and B stimuli to baseline firing and to each other, a Mann-Whitney U test was used.

Results

Behavioural analyses

Phase 1 - Appetitive conditioning. Rats were conditioned to associate an auditory cue (A) with the delivery of two sucrose pellets over 8 days (n = 7). The conditioned response was measured as response ratio which compared the number of times an animal entered the magazine during cue presentation (0-9.5 sec) to the number of magazine entries during an ITI period (1 min before cue onset). Behavioural data from this phase are presented in Figure 1b. The repeated measures ANOVA revealed a significant linear trend ($F_{(7,21)} = 3.99$, $p = .006$) with the response ratio increasing across training days. Increased response ratios suggest reward anticipatory behavior, indicating the successful acquisition of an association between the cue (conditioned stimulus [CS]) and outcome (unconditioned stimulus [US]).

Phase 2 – Appetitive & Aversive conditioning. During Phase 2, cue A continued to be reinforced with sucrose pellets. In addition, a novel auditory cue (B) was presented and reinforced with a footshock. The response ratio was calculated comparing the number of magazine entries during the cue to the average magazine entries during the inter-trial intervals (ITI). The animals were readily able to distinguish between the two cues (Fig. 1c). A repeated measures ANOVA revealed a main effect of cue ($F_{(1,19)} = 438.28$, $p < .001$), no effect of days ($F_{(3,57)} = 0.67$, $p = .573$) and no cue x day interaction ($F_{(3,57)} = 1.60$, $p = .199$).

Test. During Test sessions, the two cues were presented without reinforcement. Figure 1d shows the average response ratio to the two cues during Test. All animals were able to differentiate between the two cues seen in the lower response ratio to the cue predicting the aversive footshock compared to the cue predicting the sucrose pellets (paired t-test; $t_{(18)} = 20.44$, $p < .001$). In addition, the response ratio to cue A significantly exceeded 0.5 (paired t-test: $t_{(18)} = 7.51$, $p < .001$), indicating reward expectant behavior. The response ratio to the B

cue on the other hand fell significantly below 0.5 (paired t-test: $t_{(18)} = -41.65$, $p < .001$), indicating suppression of ongoing behavior.

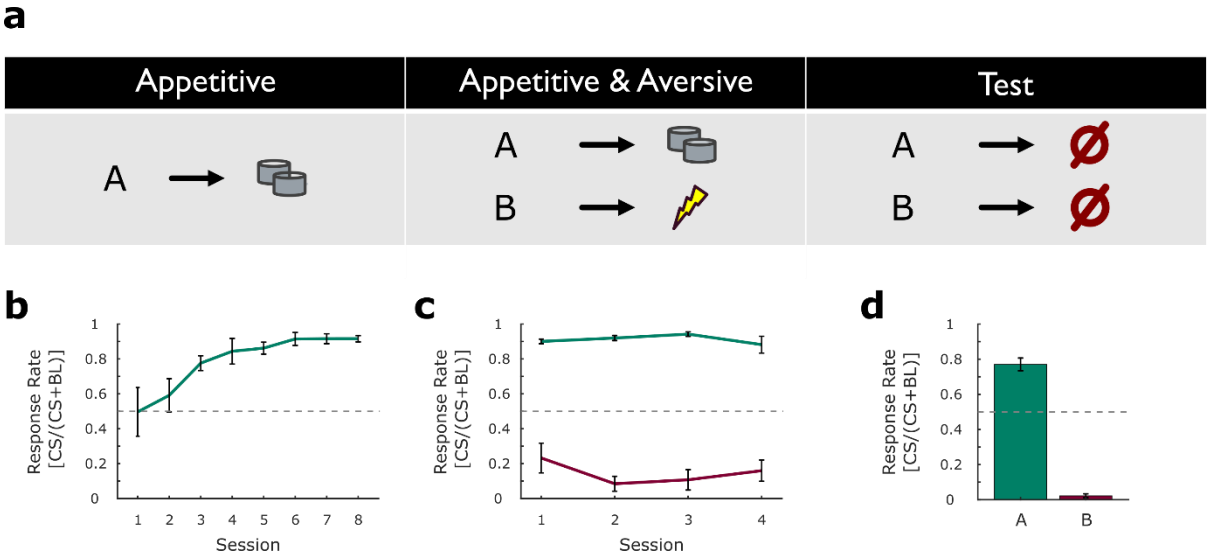


Figure 1: Conditioning paradigm and changes in conditioned response across experimental phases. Response ratio is calculated as the average percent of time spent in the magazine during the cue (0-9.5 sec) for each trial compared to the %time in the magazine during 1 min of ITI averaged over all trials of the session. **a** Schematic displaying the three phases of the conditioning paradigm. **b** Change in conditioned response across conditioning days during appetitive conditioning. Line displays response rate to cue A across conditioning days (mean ± SEM). Repeated measures ANOVA: $F_{(7,21)} = 3.99$, $p = .006$. **c** Response ratio to cues A (teal) and B (magenta) across days of appetitive and aversive conditioning. Lines display average response rate to cues A and B (mean ± SEM). Repeated measures ANOVA: cues: $F_{(1,19)} = 438.28$, $p < .001$; days: $F_{(3,57)} = 0.67$, $p = .573$; days x cues: $F_{(3,57)} = 1.6$, $p = .199$. **d** Average response rate to cue predicting reward (A; teal) and cue predicting shock (B; magenta) during Test sessions. Bar plots display the mean ± SEM. paired t-test; $t_{(18)} = 20.44$, $p < .001$.

Neural Analyses

All analyses were done on neural data collected during the Test session.

Identification of Dopamine Neurons. Dopamine neurons show characteristically wide waveforms. We utilized a waveform analysis (Takahashi et al., 2017) to identify putative dopamine neurons from VTA neurons recorded during the Test session. Out of 189 recorded VTA neurons, the cluster analysis identified 69 putative dopamine neurons (Fig. 2b). These neurons displayed a significantly smaller amplitude ratio and longer spike duration (Two-sample t-test; amplitude ratio: $t_{181} = -14.35$, $p < .001$; spike duration: $t_{181} = 16.53$, $p < .001$).

Reward Prediction Error (RPE) in VTA dopamine neurons. As a first step we analyzed the firing patterns of the putative dopamine neurons to the reward predicting trials. We observed the well-documented RPE signal: Phasic excitation at cue A onset (median = 1.61; Mann-Whitney-U test: U = 5.46, $p < .001$), phasic inhibition at reward omission (Fig. 3a; median = -0.54; Mann-Whitney-U test: U = -4.31, $p < .001$) and a negative correlation between these epochs (Fig. 3b; Spearman correlation: $r = -0.49$, $p < .001$). The higher the FR at cue onset, the lower the FR at US omission, suggesting the presence of a reward prediction error and confirming findings of previous studies (Matsumoto et al., 2016; Schultz et al., 1997; Takahashi et al., 2011). The inverse of the RPE signal observed to the A cue was not observed to cue B. At the onset of the B cue, we observed phasic excitation of average population firing above baseline activity (median = 1.21; Mann-Whitney-U test: U = 5.90, $p < .001$), while average activity at shock omission was not significantly different from baseline (median = 0.13; Mann-Whitney-U test: U = 1.22, $p = 0.221$). A comparison between cue B onset and shock omission revealed no correlation (Fig. 3d; Spearman correlation: $r = 0.10$, $p = .422$). These results confirm a well-established RPE signal but show that an aversive PE (APE) is not immediately clear.

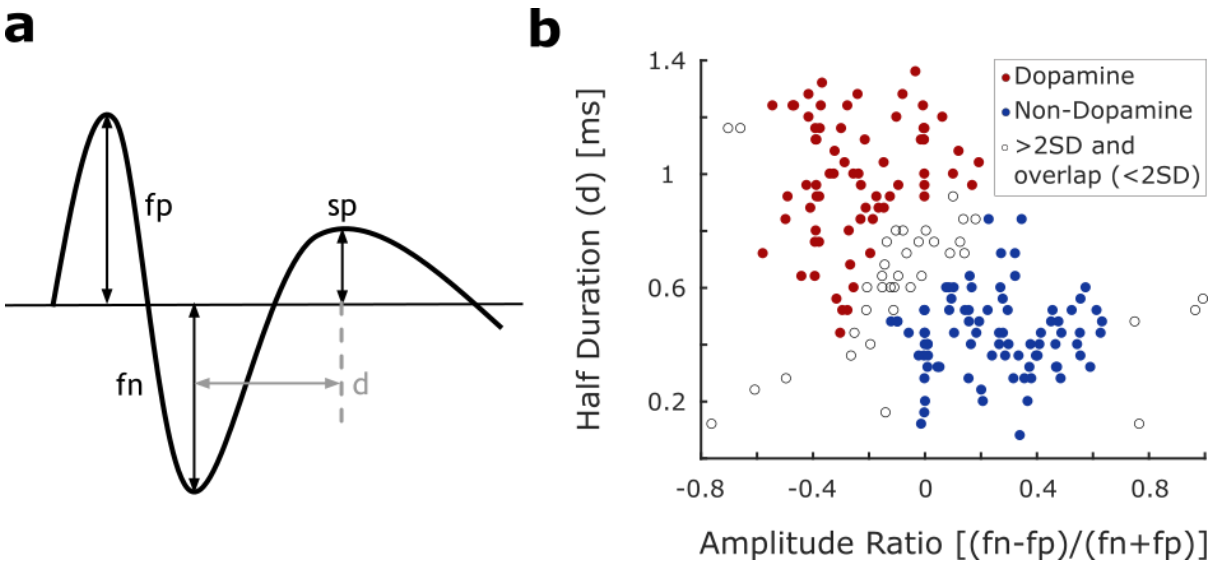


Figure 2: Classification of VTA neurons based on waveform characteristics. **a** Schematic of a waveform. Amplitude values of the first positive (fp) and first negative (fn) segments are used to calculate the amplitude ratio. Half duration (d) describes the time between fn and second positive segment (sp). **b** Cluster analysis was conducted using the two waveform characteristics amplitude ratio and half duration to classify dopamine (red) and non-dopamine (blue) neurons. Only neurons within 2 s.d. of one cluster's centroid were considered as classified. Neurons that fell within 2 s.d. of both clusters or were more than 2 s.d. away from at least one centroid were not considered and are displayed as open circles.

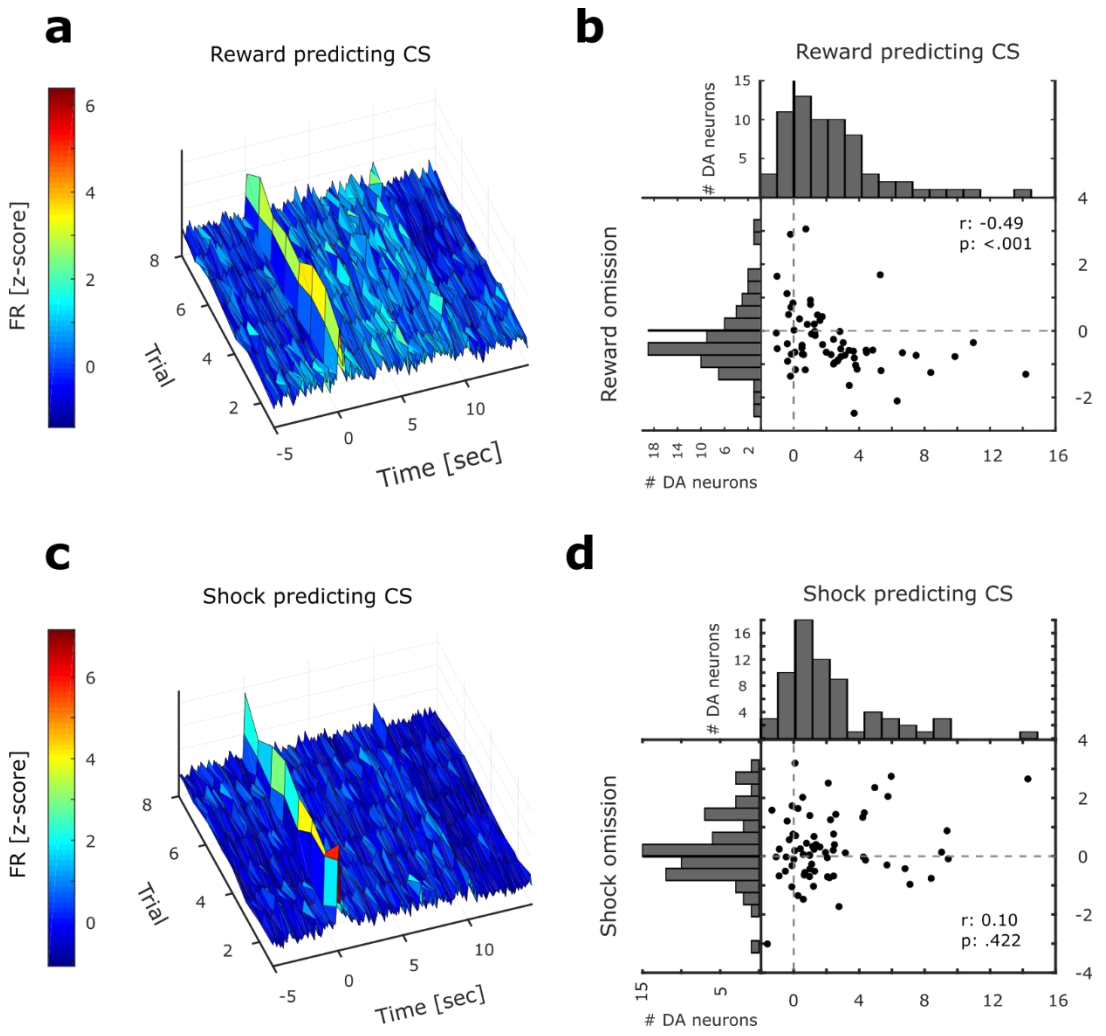


Figure 3: Comparison of firing rates to cues A and B. **a** 3D heatmap of average normalized FR to cue A across trials. Time axis is aligned to cue onset, with a 5 sec pre-cue period being displayed. Cue presentation ranges from 0-10 sec, followed by the outcome omission period. FR is normalized using a z-score based on the distribution of inter-trial interval activity. **b** Correlation of DA activity comparing reward predicting cue A (0-500ms) to time of reward omission. (Spearman correlation: r = -0.49, p < .001) **c** 3D heatmap of average normalized FR to cue B across trials. FR is normalized to activity from the inter-trial interval. The cue period ranges from 0-10 sec, followed by the outcome omission period. **d** Correlation of DA activity comparing shock predicting cue B (0-500ms) to time of shock omission (Spearman correlation: r = 0.10, p = .422).

Magnitude, duration, and timing of modulation do not differ between cues. To understand whether and how cues predicting appetitive and aversive events are distinguished by midbrain DA neurons, we compared neural activity between cues A and B on a population and single neuron level. A two-sample Kolmogorov-Smirnoff test revealed that while the distributions of normalized firing rates differ significantly between reward omission and shock omission (Fig. 4b; two-sample KS test: $k = 0.33$, $p = .001$), there is no difference between dopamine activity at the onset of A compared to B (Fig. 4a; two-sample KS test: $k = 0.13$, $p = .55$) on a population level. Even though the average population activity does not seem to differ between A and B, individual neurons could still differentiate between the cues. We therefore calculated the absolute difference in normalized FR between A and B trials for individual neurons across baseline, cue, and outcome omission periods, we observed that neurons respond most differently during the first second of cue presentation and the first 500ms of outcome omission (Fig. 4d).

We examined the magnitude, duration, or timing of normalized FR to A and B in order to determine whether these modalities of DA activity could underly the animals' ability to discriminate between cues of opposite valance. For this, we first identified the maximum and minimum normalized FR for each neuron throughout the 10 sec of cue presentation. Analysis of the empirical cumulative density function (eCDF) showed no significant difference in distribution of maximum (Fig. 5a; Two-sample KS test: $k = 0.13$, $p = .57$) and minimum (Fig. 5b; Two-sample KS test: $k = 0.09$, $p = .95$) normalized FR between the two cues. In addition, we identified the exact time at which maximum and minimum firing occurred. A Cox regression determined that the majority of DA neurons were maximally modulated during the first second of cue presentation. However, there was no significant difference in timing of maximum (Fig. 5c; Cox regression: $e^b = 1.00$, $p = .18$) or minimum FR (Fig. 5c; Cox regression: $e^b = 1.00$, $p = .90$) between A and B trials.

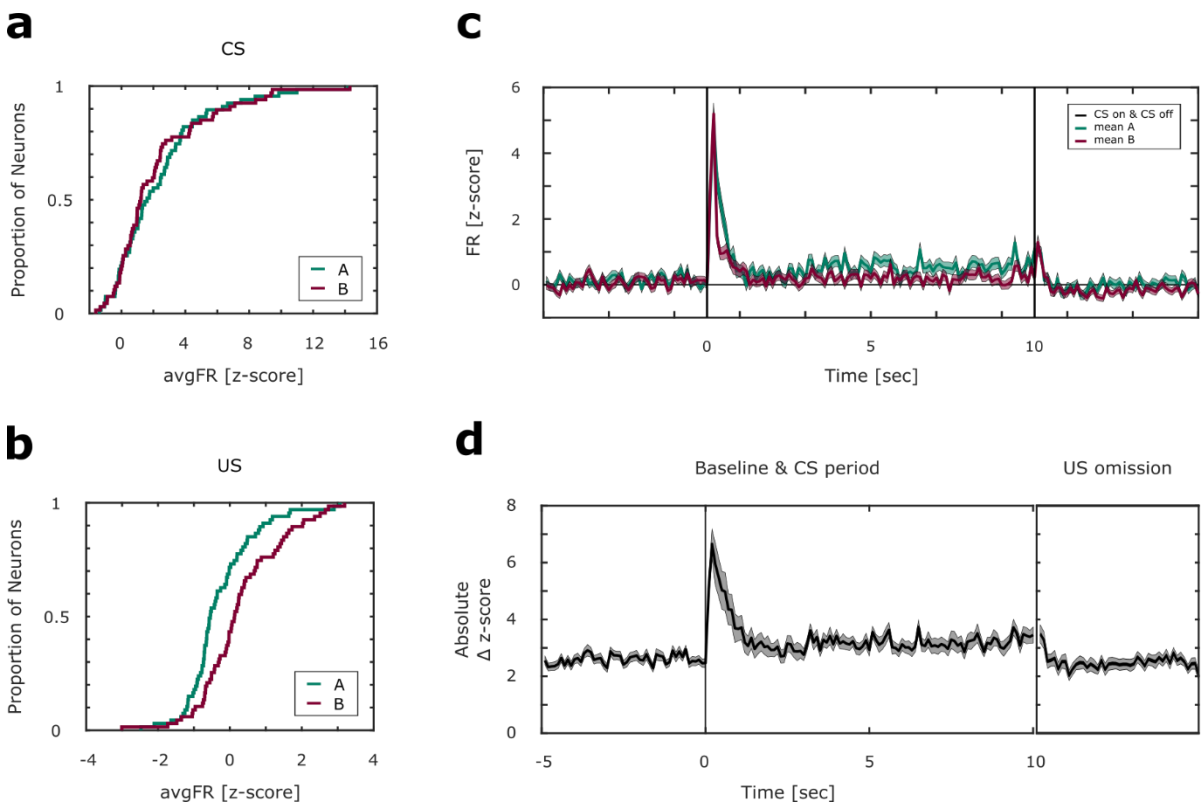


Figure 4: Comparison of firing rates to reward and shock predicting cues. **a** Empirical cumulative density function (eCDF) of the average normalized FR to the onset (0-500ms) of reward predicting cue A (teal) and shock predicting cue B (magenta). Two-sample KS test: $k = 0.13$, $p = .550$. **b** eCDF of the average normalized FR to A (teal) and B (magenta) at time of outcome omission. Two-sample KS test: $k = 0.33$, $p = .001$. **c** Peristimulus time histogram of population normalized FR throughout baseline, cue and outcome omission periods (mean ± SEM). cue onset and offset are displayed as vertical, black lines at 0 and 10 sec. **d** By-neuron absolute difference in normalized FR throughout baseline and cue period (mean ± SEM). Period of outcome omission is aligned to time of expected outcome offset.

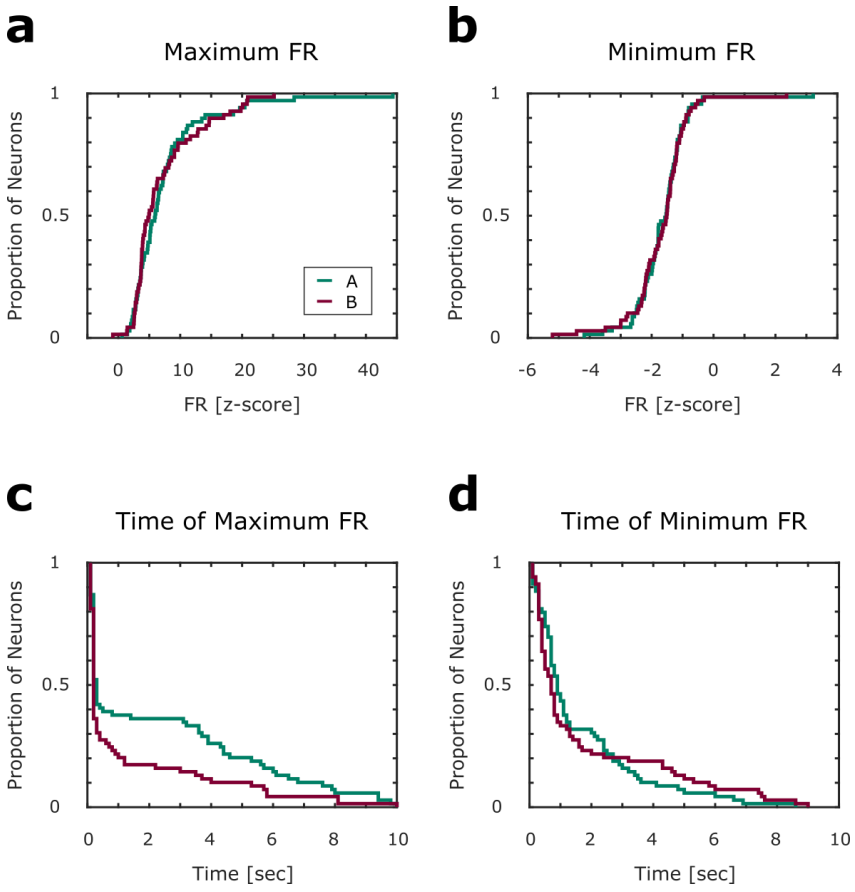


Figure 5: Magnitude and timing of maximal modulation during A (teal) and B (magenta) trials. **a** Distribution of maximum FR observed during cue presentations. Two-sample KS test: $k = 0.13$, $p = .57$. **b** Distribution of minimum FR observed during cue presentations. Two-sample KS test: $k = 0.09$, $p = .95$. **c** Time during cue presentation at which maximum FR was observed as function of proportion of DA neurons. Cox regression: $e^b = 1.00$, $p = .18$. **d** Time at which minimum FR was observed during cue presentation as function of proportion of DA neurons. Cox regression: $e^b = 1.00$, $p = .90$.

Next, we identified periods during which dopamine activity was significantly modulated above or below baseline. Using the cumulative sum of FRs to each cue and baseline period (see Methods), we determined periods of excitation and inhibition for each neuron. Adding all periods that were identified in such a way, we calculated the overall duration of excitation and inhibition by neuron. Comparing the total duration of significant modulation between reward- and shock-predicting cues, we found no difference between the two cues for either duration of excitation (Fig. 6a; two-sample KS test: $k = 0.19$, $p = .15$) or inhibition (Fig. 6b; two-sample KS test: $k = 0.16$, $p = .32$). Most dopamine neurons that showed excitation at some point during the presentation of the reward- or shock-predicting cue, were first excited during the first half of the cue presentation. The Cox regression revealed that the onset of excitation occurred at similar times to A as to B (Fig. 6c; $e^b = 1.00$, $p = .90$). The same was true for periods of inhibition. The majority of dopamine neurons showing inhibition were inhibited during the first half of cue presentation, for A and B (Fig. 6d; Cox regression: $e^b = 1.02$, $p = .60$).

These data suggest that modulation occurs early during both cues. However, neither magnitude nor timing of maximal modulation of DA firing underlie the distinction of reward and aversion predicting cues. In addition, total duration and onset of significant modulation does not differ between A and B.

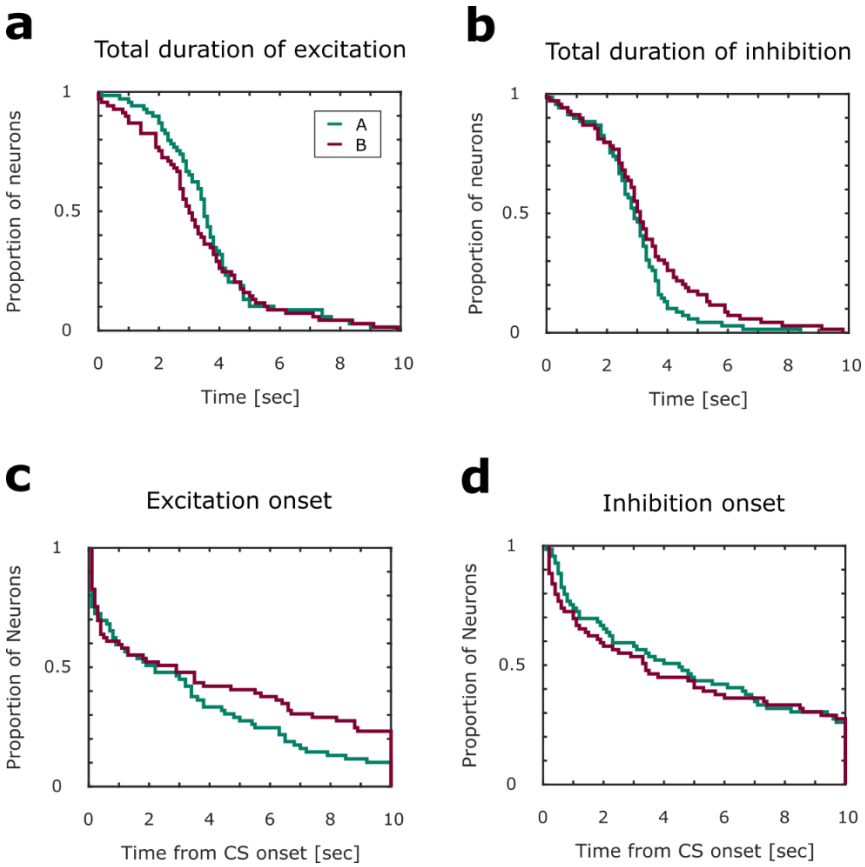


Figure 6: Duration and onset of significant excitatory and inhibitory periods to cue A (teal) and B (magenta). **a** eCDF of the total duration of excitation throughout the cue period in seconds. Two-sample KS test: $k = 0.19$, $p = .15$. **b** eCDF of the total duration of inhibition throughout the cue period in seconds. Two-sample KS test: $k = 0.16$, $p = .32$. **c** Onset of excitation across the cue period as a function of proportion of DA neurons. Cox regression: $e^b = 1.00$, $p = .90$. **d** Onset of inhibition across cue periods as a function of proportion of DA neurons. Cox regression: $e^b = 1.02$, $p = .60$.

Dopamine neurons can be divided into subpopulations based on activity patterns.

Next, we identified the overall proportion of neurons that were phasically modulated to A and B. Neurons were segregated according to their responses to at onset of A and B, which showed that within the DA population, 45% of neurons only showed excitation during the start of A trials., 28% showed only inhibition, and 16% were both exited and inhibited, and 12% were not modulated. By comparison, 35% of neurons were phasically excited, 33% inhibited, 13% were both excited and inhibited, and 19% were not modulated by B (Supplemental Fig. 1a; Chi-square test of independence: $X^2 = 2.66$, $p = .457$). Next, we focused on the neurons that showed one type of modulation (excitation or inhibition, but not both) and examined whether and how many of these neurons displayed the same or opposing activity patterns during the start of A and B. Out of 69 dopamine neurons, 12 were significantly excited, 3 neurons were significantly inhibited to both A and B, 9 were excited to A and inhibited to B, and 7 were inhibited to A and excited to B. That is, 21% of neurons were modulated in the same direction, and 23% in the opposing direction to A and B (Supplemental Fig. 1b; Fisher's exact test: *Odds Ratio* = 0.57, 95% CI [0.115, 2.845], $p = .697$). Neurons that fall into the second group, exhibiting excitation to one cue and inhibition to the cue of opposite valence, fit with the integrated appetitive-aversive systems model.

Cue activity accurately represents cue identity. To determine whether midbrain dopamine neurons discriminate between appetitive and aversive events at the pseudo ensemble level, we used two classification algorithms, a logistic generalized linear model (GLM) and a linear discriminant analysis (LDA). We used a principal component analysis (PCA) on the FR segregated into 100ms epochs throughout A and B (Supplemental Fig. 2a). The GLM and LDA models were trained on 50% of A and B trials of one 100ms epoch and tested on the remaining left-out trials from the same epoch (see Methods for details). This was done for every epoch throughout the 10sec of cue presentation. A random model was

created by using the same analyses on shuffled baseline data. Cue identity was best decoded at cue onset (Fig. 7a&c). During the first second of cue presentation, more than 80% of trials were correctly classified by both GLM and LDA models ($\hat{\mu}_{decoder} > \hat{\mu}_{random} \pm 95\%$ CI; GLM: avg. 95% CI: [0.319; 0.703]; $Accuracy_{min >CI} = 0.706$; LDA: avg. 95% CI: [0.290; 0.719]; $Accuracy_{min >CI} = 0.726$; no epochs fell below the lower CI bound of the respective model). Decoding accuracy at later time points during the cue period was similar to the random model and did not exceed the 95% CI. The confusion matrixes show that this effect is not driven by one particular trial type. Analyzing the average classification accuracy of the models during the first second of cue presentation, we saw that 82% of A trials are correctly classified as the reward predicting cue by the GLM, with 18% misclassified trials. Similarly, 92% of trials predicting shock were correctly identified as B trials, with only 8% misclassification rate (Fig. 7b). The LDA achieved 78% correct classification of reward predicting trials during the first second of cue presentation, while shock predicting trials were 80% correctly classified. Only 22% of A trials and 20% of B trials were misclassified by the LDA (Fig. 7d). These data suggest that information about cue identity is contained primarily in DA activity at cue onset, corresponding with phasic activity patterns. However, it does not provide information about whether this phasic activity represents valence.

Activity at outcome omission misclassifies cue activity as events of opposite valence. To identify whether information about valence is coded by midbrain DA neurons at the time of outcome omission, we conducted a cross-temporal analysis, comparing FRs at time of outcome omission to activity throughout the cue period. We conducted a PCA on the period of reward and shock omission (Supplemental Fig. 2b) and trained the GLM and LDA models on FRs from the same period. The two models were then tested on activity from 100ms epochs throughout the cue period. Surprisingly, decoder performance of both GLM and LDA models was significantly worse than the random model at the start of the cue ($\hat{\mu}_{decoder} < \hat{\mu}_{random}$

± 95% CI). Following this initial period, decoder performance flipped and exhibited periods of significantly better performance than the random model, which disappeared towards the latter half of cue presentation (Fig. 7e&g; $\hat{\mu}_{decoder} < \hat{\mu}_{random} \pm 95\%$ CI; GLM: 95% CI: [0.304; 0.726]; $Accuracy_{min\ <CI} = 0.202$; $Accuracy_{min\ >CI} = 0.743$; LDA: 95% CI: [0.307; 0.731]; $Accuracy_{min\ <CI} = 0.232$; $Accuracy_{min\ >CI} = 0.793$). By analyzing the average accuracy of the outcome decoders during the first 500ms of cue presentation, we observed that 71% of A trials and 72% of B trials were misclassified by the GLM decoder during the first 500ms of the cues. Only 29% of reward predicting and 28% of shock predicting trials were accurately classified as such by activity from US omission. Similarly, 68% of A trials and 71% of B trials were misclassified by the LDA model, while only 32% of reward predicting and 29% of shock predicting trials were correctly identified as such during at the start of cue presentation (Fig. 7f&h). It appears as if opposite information is contained in DA activity during the first 500ms of reward and shock predicting cues compared to reward and shock omission.

This suggests that on a population level, valence information about a predicted event is represented by phasic DA firing at cue onset, and omission of this outcome is represented as event of opposite valence. However, the reason for the switch in accurate decoding, leading to better performance compared to a random model remains unclear. Based on extensive research into the role of VTA DA in associative learning (Cohen et al., 2012; Tobler et al., 2005; Wang & Tsien, 2011), one possible explanation could be that there exists a valance prediction error signal represented by a signed change in phasic firing at cue onset, followed by an unsigned prediction signal (see Maes et al., 2020). Therefore, we examined whether this bimodal signal seen in the population is represented in individual neurons, or in different neurons resulting in two distinct populations.

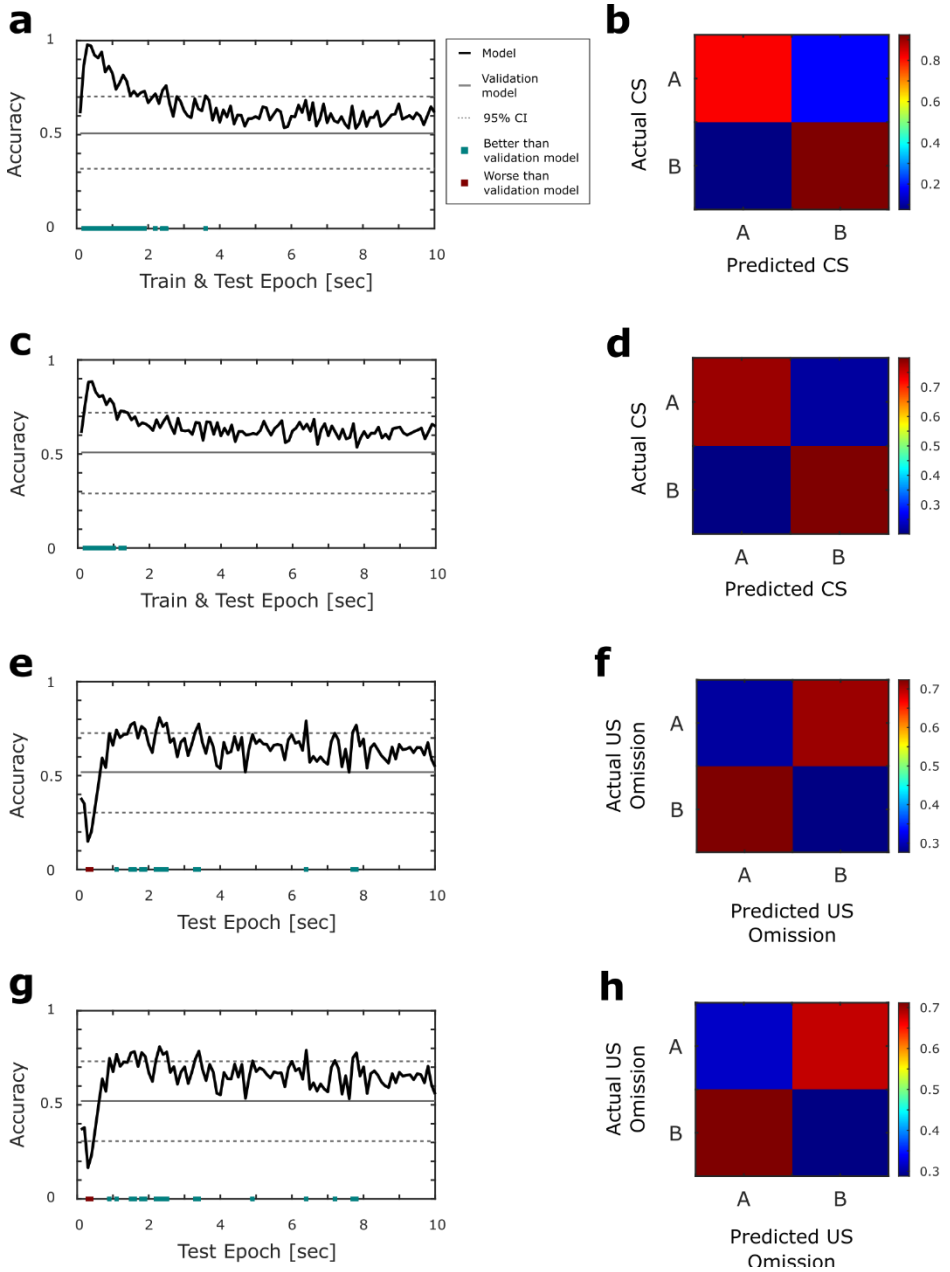


Figure 7: Dopamine neurons decode cue identity from firing during the cue and at outcome omission. **a** Generalized linear model (GLM) decoding cue identity from cue firing, comparing model performance to a random model (95% CI: [0.319; 0.703]; $Accuracy_{min >CI} = 0.706$). **b** Confusion matrix of GLM cue decoder indicating percent of correctly and incorrectly classified A and B trials based on decoder accuracy of epochs between 0-1sec. **c** Linear discriminant analysis (LDA) decoding cue identity from cue firing. Cue decoder accuracy exceeds that of a random model at cue onset (95% CI: [0.290; 0.719]; $Accuracy_{min >CI} = 0.726$). **d** Confusion matrix of cue decoder LDA indicating percent of correctly and incorrectly classified A and B trials based on average decoding accuracy of the epochs between 0-1 sec. *(Figure legend continues on the next page)*

(Continuing legend; Figure 7) **e** GLM outcome decoder accuracy when model trained on outcome omission FR is tested across cue epoch FR (95% CI: [0.304; 0.726]; $Accuracy_{min\ <CI}$ = 0.202; $Accuracy_{min\ >CI}$ = 0.743). **f** Confusion matrix of outcome decoder GLM during the first 500ms of cue presentation indicating the percent of correctly and incorrectly classified trials. **g** Cross-temporal LDA trained on FR at outcome omission and tested on FR of cue epochs. Accuracy of the model at each cue epoch is indicated as black line (95% CI: [0.307; 0.731]; $Accuracy_{min\ <CI}$ = 0.232; $Accuracy_{min\ >CI}$ = 0.793. **h** Confusion matrix of LDA outcome decoder performance during epochs between 0-500ms of cue presentation indicating correctly and incorrectly classified A and B trials. Black lines indicate average accuracy of models, grey lines indicate validation models based on baseline firing. Dotted lines indicate the 95% CI of validation models. Epochs of better than random performance are indicated as blue squares, epochs of worse than random performance are indicated as red squares.

Separate populations of DA neurons code a valence prediction error and prediction.

To identify possible populations underlying the switch in decoder performance, we used a by-neuron cross-temporal classifier model. This GLM was trained on activity during outcome omission of each neuron and tested on cue epoch activity of the same neuron. We then compared the accuracy for each neuron to a validation model which was tested on randomized baseline activity. Based on earlier results from the population decoder, we saw that on a population level, decoder accuracy was worse than a random model between 300-800ms after cue onset. We therefore investigated whether decoding accuracy of individual neurons was better or worse than accuracy of the random by-neuron model between 300-800ms after cue onset. Interestingly, we found two separate populations that show distinct decoding as well as distinct firing patterns throughout the cue. One population performed better than a random model not just at cue onset, but throughout the whole cue period (Fig. 8a; n=28; neuron with smallest difference to its 95% CI: [19.00%; 53.19%], Accuracy: 62.50%). This population displays more sustained activity during cue presentation (Fig. 8b; Mann-Whitney-; A vs. BL: U_{min} = -1.99; p = .046; B vs. BL: U_{min} = -1.97, p = .049) and

during the outcome omission period (Mann-Whitney-; A- vs. BL: U_{min} = -1.97; p = .049; B- vs. BL: U_{min} = -1.97, p = .049). Calculating the absolute difference in z-scores between the two trial types revealed slight but sustained differences throughout the cue period and very minimal differences between activity to reward and shock omission (Fig. 8c). Firing rates did not differ between the reward predicting and shock predicting cue (300-800ms; Fig. 8i; Mann-Whitney U: U = -1.02, p = .306). Interestingly, also firing rates at time of outcome omission were not different between reward- (A-) and shock-omission (B-; Fig. 8j; Mann-Whitney-U: U = -1.01, p = .314). Comparing the average FR of the two cues at cue onset to the time of omission of the incongruent outcome (A to B- and B to A-), the two events were not significantly correlated (Fig. 8d left; Spearman correlation: r = -0.16, p = .235). Only when comparing cue onset activity to the congruent outcome omission (A to A- and B to B-) revealed a positive relationship (Fig. 8d right; Spearman correlation: r = 0.28, p = .040).

The second population on the other hand, performed worse than a random model at cue onset only and decoded the cues at the level of a random model throughout the rest of cue presentation (Fig. 8e; n=18; neuron with smallest difference to its 95% CI: [28.13%; 75.69%], Accuracy: 25.00%). These DA neurons exhibited firing patterns that coincide with a valence prediction error signal. Between 300-800ms after cue onset, neurons from this population responded with significantly different firing rates to A and B cues (Fig. 8k; Mann-Whitney-U: U = 3.88, p < .001). Neurons in this population showed phasic excitation to A (Mann-Whitney-U: A vs. BL: U_{min} = 2.01, p = .045) and inhibition to B (Fig. 8f; Mann-Whitney-U: B vs. BL: U_{min} = -1.98, p = .048). At time of outcome omission, these neurons were phasically inhibited to reward omission (Mann-Whitney-U; A- vs. BL: U_{min} = -2.04, p = 0.041) and phasically excited to shock omission (Mann-Whitney-U; B- vs. BL: U_{min} = -2.04, p = .041). Firing rates to reward and shock omission differed from each other (Fig. 8l; Mann-Whitney-U; A- vs. B-: U = -4.57, p < .001). This was also reflected by the absolute difference

in z-scores revealing pronounced, phasic differences between A and B cues, as well as reward and shock omission, while the remainder of the cue period showed differences comparable to baseline (Fig. 8g). As expected in a prediction error population, activity at time of cue onset (300-800ms) was negatively correlated with incongruent (A to B- and B to A-) activity at time of outcome omission (Fig. 8h left; Spearman correlation: $r = 0.60$, $p < .001$). Interestingly, cue activity was also positively correlated with congruent activity (A to A- and B to B-) at time of outcome omission (Fig. 8h; Spearman correlation: $r = -0.53$, $p = .001$).

To summarize, we found two separate populations of VTA DA neurons that exhibited distinctly different activity patterns to reward and shock predicting cues, as well as omission of both reward and shock. The population that displayed similar activity patterns between A and B correctly decoded cue identity based on activity from the outcome omission period. While this makes cue decoding based on valence by this population unlikely, we do not have sufficient evidence for an informed hypothesis about this subpopulation of VTA DA neurons. On the other hand, we found a second population that exhibited opposite, phasic activity patterns between reward and shock predicting cues, as well as between reward and shock omission. These DA neurons decode activity at cue onset as the opposite cue, based on activity from the outcome omission period, in line with a valence-based prediction error.

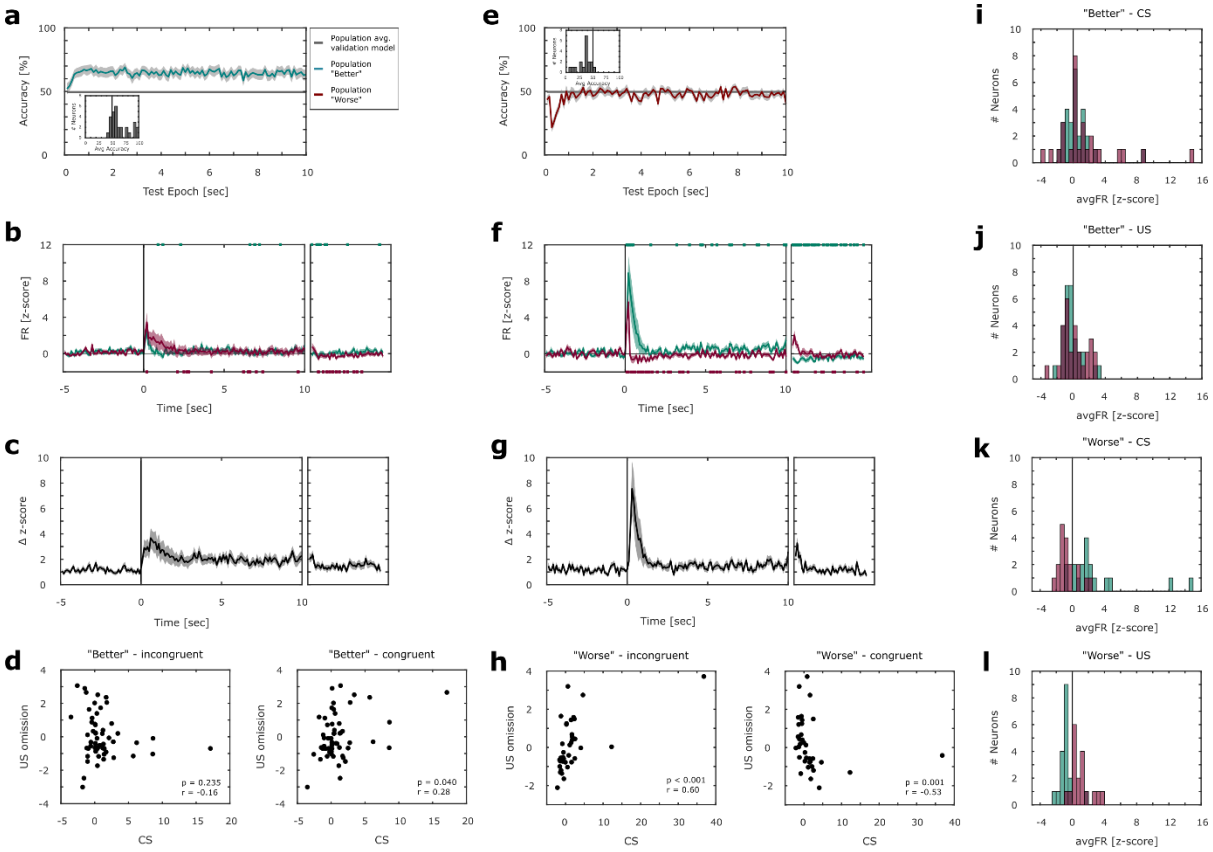


Figure 8: Two distinct populations of midbrain DA neurons exhibit different decoder and activity profiles. **a** Average accuracy ± SEM of the population decoding cue identity better (population "Better" - blue line) than a validation model (dark grey line) at 300-800ms after cue onset based on FR from the outcome omission period. Inlay displays a histogram of the average accuracy of each neuron in this population between 300-800ms. **b** PSTH of normalized FR of population "Better" (n = 28) throughout the second half of baseline FR (BL; -10 – 0 sec), presentation of A (teal) and B (magenta), as well as reward (A-) and shock omission (B-). Lines represent mean ± SEM. Epochs which are significantly different from BL FR are indicated as squares on the top (A cue and reward omission; teal) and bottom (B cue and shock omission; magenta) x axes. Mann-Whitney-U test; A vs. BL: U_{min} = -1.99, p = .046; B vs. BL: U_{min} = -1.97, p = .049; A- vs. BL: U_{min} = -1.97, p = .049; B- vs. BL: U_{min} = -1.97, p = .049. **c** Absolute difference in z-scores throughout baseline, cue and outcome omission period of "Better" subpopulation. Line displays mean ± SEM. **d** Average normalized cue FR (300-800ms) compared to incongruent (left; A vs B- and B vs. A-) and congruent (right; A vs. A- and B vs. B-) outcome omission FR of neurons from subpopulation "Better". Spearman correlation: Incongruent: r = -0.16, p = .235; Congruent: r = 0.28, p = .040. **e** Population of neurons decoding cue identity worse (population "Worse" - red line) than a validation model (dark grey line) at cue onset (300-800ms) based on FR from the outcome omission period. Average accuracy is plotted as mean ± SEM. Inlay represents a histogram of average accuracy of each neuron in this population during 300-800ms. *(Figure legend continues on next page)*

(Continuing legend; Figure 8)

f PSTH of population "Worse" (n = 18) during baseline FR (BL; -10 – 0 sec), presentation of A (teal) and B (magenta), as well as reward (A-) and shock omission (B-). Lines display mean ± SEM. Epochs which are significantly different from BL FR are indicated as squares on the top (A cue and A-; teal) and bottom (B cue and B-; magenta) x axes. Mann-Whitney-U test; A vs. BL: U_{min} = 2.01, p = .045; B vs. BL: U_{min} = -1.98, p = .048; A- vs. BL: U_{min} = -2.04, p = .041; B- vs. BL: U_{min} = -2.04, p = .041. **g** Absolute difference in z-scores throughout baseline, cue and outcome omission period of "Worse" subpopulation. Line represents mean ± SEM. **h** Average normalized cue FR (300-800ms) compared to incongruent (left; A vs B- and B vs. A-) and congruent (right; A vs. A- and B vs. B-) outcome omission FR of neurons from subpopulation "Worse". Spearman correlation: Incongruent: r = 0.60, p < .001; Congruent: r = -0.53, p = .001. **i-l** Histograms of average normalized FR of neurons to cue A and reward omission (A-; teal) as well as cue B and shock omission (B-; magenta). **i** Average normalized FR of population "Better" during 300-800ms after cue onset. Mann-Whitney-U: A vs. B: U = -0.62, p = .537. **j** Average normalized FR during outcome omission of population "Better". Mann-Whitney-U test; A- vs. B-: U = -0.33, p = .740. **k** Histogram of average normalized FR of population "Worse" during cue onset (300-800ms). Mann-Whitney-U: A vs. B: U = 3.98, p < .001. **l** Distribution of average normalized FR during outcome omission of population "Worse". Mann-Whitney-U: A- vs. B-: U = -3.60, p < .001.

Discussion and Future Directions

The current study investigated whether activity patterns of dopamine (DA) neurons in the ventral tegmental area (VTA) represent a valence-based prediction error in the brain. Further, we wanted to understand if the reason previous studies found conflicting results when researching DA activity during aversion was due to DA subpopulations representing opposing appetitive and aversive systems. We therefore analyzed activity patterns of midbrain DA neurons during cues predicting appetitive and aversive events as well as during omission of these expected events. Our data showed modulation to both appetitive and aversive cues, however, magnitude, duration and timing of the modulation was not different between the opposing cues. On a population level, we found that cue identity is decoded by DA neurons only at the start of cue presentation, and outcome omission is misclassified as event of opposite valence during the same cue period. So far, these findings support both the prediction error and opposing systems theories. However, we identified two subpopulations of VTA DA neurons that behave differently. One group of neurons correctly identified cue identity throughout the whole cue period based on activity at outcome omission and did not show differences in modulation between reward and aversion predicting cues or reward and shock omission. This suggests a valence-free prediction that is carried through the whole period of cue presentation. On the other hand, the second population of DA neurons showed valence decoding only at cue onset and exhibited bidirectional phasic activity at the start of cue presentation and at outcome omission. This population's activity, excitation to reward predicting cues and shock omission and inhibition to shock predicting cues and reward omission, is in line with a valence-based PE theory. Interestingly, we did not find evidence for a third population representing an aversive system, decoding valence-based information through excitation to aversive events and inhibition to rewarding events. Together, our data

suggest that not all, but a subpopulation of midbrain DA neurons support learning by coding a valence-based prediction error.

Ventral tegmental area dopamine – key player in reinforcement learning. The ventral tegmental area (VTA) and more specifically phasic dopamine (DA) activity in the VTA have been implicated to play a key role in reinforcement learning and motivation (Bayer et al., 2007; Chang et al., 2016; Maes et al., 2020; Steinberg et al., 2013; Waelti et al., 2001). This is not surprising given the various projection targets of VTA DA neurons, most notably the nucleus accumbens (NAcc) and medial prefrontal cortex (mPFC) (Lammel et al., 2011, 2014). Particularly DA inputs to the NAcc are necessary for motivational, goal-directed, and drug-seeking behaviors (Cheer et al., 2007; Mannella et al., 2013; Robbins & Everitt, 1996; Yun, 2004), while inputs to the mPFC are implicated in decision making and promote avoidance in fear learning (Vander Weele et al., 2018). However, goal-directed behaviors and decisions do not only entail behavior towards rewarding outcomes, but also avoiding negative outcomes. Therefore, VTA DA activity is thought to play a role not only in learning about rewards, but also aversive events (Brischoux et al., 2009; Cohen et al., 2012; Lammel et al., 2012; H. Matsumoto et al., 2016; M. Matsumoto & Hikosaka, 2009b; Zweifel et al., 2011). With such widespread projections of DA neurons to key regions, a valence-based prediction error (PE) signal would be the most economic and clear form of transmitting valence information across the brain to drive decisions and behaviors that ensure survival. However, this is only true if the receiving region requires an influx of DA elicited by excitation of DA neurons in the VTA to rewarding events and a decrease in DA elicited by inhibition of DA neurons in the VTA to aversive events in order to elicit the appropriate behavioral response. Another possibility would be that some target regions require an influx of DA to appropriately avoid aversive events. This would require two separate populations of neurons, one encoding aversive information through inhibition and the other through excitation. The

field has found diverse firing patterns of VTA DA neurons to aversive events (Brischoux et al., 2009; Cohen et al., 2012; H. Matsumoto et al., 2016; M. Matsumoto & Hikosaka, 2009b; Salinas-Hernández et al., 2018; Wang & Tsien, 2011; Zweifel et al., 2011). We therefore wanted to gather information on whether midbrain DA neurons show activity patterns in line with a valence-based prediction error, in addition to understanding whether these contradicting findings could be due to subpopulations of DA neurons representing opposing aversive and appetitive systems.

Prediction error in midbrain dopamine neurons. It has been shown many times that dopamine neurons signal the reward PE by bidirectional modulation of firing rate during reward learning (Cohen et al., 2012; Fiorillo et al., 2008; Schultz et al., 1997; Waelti et al., 2001; Wang & Tsien, 2011). Most studies using *in-vivo* electrophysiology focus on a 500ms-1sec time window after cue onset, as this has been shown to capture the phasic response during reward learning (H. Matsumoto et al., 2016; Schultz, 2007; Takahashi et al., 2017). Utilizing this information from previous studies, we compared DA activity during the first 500ms of cue presentation to activity during a window at outcome omission of the same size. We found the reward predicting cue and reward omission to be negatively correlated. This suggests the presence of a reward PE in the recorded population of midbrain DA neurons, as shown by previous studies.

However, if dopamine neurons signal a valence-based PE, one would expect to find a reverse response to aversive cues. An unexpected aversive event would be predicted to cause an inhibition in firing, compared to excitation to reward. Additionally, the omission of an anticipated aversive event would be predicted to cause excitation compared to inhibition observed at the omission of an expected reward. Some studies found evidence supporting a valence PE hypothesis by observing inhibition to aversive events (H. Matsumoto et al., 2016; Mileykovskiy & Morales, 2011; Ungless et al., 2004) or excitation to the omission of an

aversive event (H. Matsumoto et al., 2016; Salinas-Hernández et al., 2018), others claimed that midbrain DA neurons only code reward PE (Fiorillo, 2013; Mirenowicz & Schultz, 1996). In this study, comparing DA activity during the onset of the shock predicting cue to activity during shock omission, the two events were not related in a negative fashion. Instead, we found a non-significant positive correlation. At first glance, this would suggest the absence of a valence-based PE.

Differentiation between reward and aversion predicting cues. DA neurons could still drive learning in general, as observed by differential behavioral output to appetitive and aversive cues. For this, DA activity patterns would need to distinguish between these cues, but the firing patterns would not have to represent a PE signal. Two studies suggested that the magnitude of modulation could underly the differentiation between cues (Joshua et al., 2008; Tobler et al., 2005). Similarly, timing of activity could provide a way to signal differing cues. To understand whether and how exactly DA neurons differentiate between events and cues predicting these events requires analysis of not just a brief period of 500ms-1sec, but a detailed investigation of different time points throughout the cue presentation. Analyzing the whole duration of cue presentation, we found evidence that VTA DA neurons are modulated by both appetitive and aversive events and cues predicting these events. Importantly, magnitude, duration, and timing of modulation did not differ between reward and shock predicting cues. This suggests that a different mechanism must underly the differentiation of cues.

Another option would be that different brain regions support learning about events of opposite valence, with the ventral tegmental area only being involved in reward learning and not representing valence information. In this case however, midbrain DA neurons should still show different activation to appetitive and aversive cues – modulation to appetitive events and no modulation to aversive events. Only in the case that the VTA DA signal is only

utilized to form associations between cues and predicted events, with other brain regions and neural populations providing valence information, would the activity pattern not differ between opposing cues. In this case, DA neurons would not distinguish between cues. Along with previous studies (H. Matsumoto et al., 2016; M. Matsumoto & Hikosaka, 2009b; Takahashi et al., 2011; Wang & Tsien, 2011), we found clear evidence that DA neurons in the VTA distinguish between two opposing cues, correctly identifying trial type. We identified FRs to be most different at cue onset and at outcome omission between reward and shock predicting cues. In addition, two separate classifier models showed that the trial types were very well identified by neural firing. However, this was only true early in cue presentation. Later during the cue, classifier accuracy did not exceed that of a random model, hovering around 50% correct/incorrect identification.

VTA DA neurons represent valence information of both appetitive and aversive events. We also found evidence that midbrain DA neurons do indeed represent valence. By comparing activity during reward and shock omission to activity during the cues presenting these events, we found that outcome omission firing coincided with activity to the cue predicting the opposite outcome. Two separate models classified cue firing as the opposite cue when trained on activity from outcome omission, suggesting reward predicting cues and reward omission just as shock predicting cues and shock omission represent opposing valences. This would be in line with both a valence PE theory and a hypothesis of two opposing systems. Additionally, our results support findings of previous studies (Guarraci & Kapp, 1999; H. Matsumoto et al., 2016; M. Matsumoto & Hikosaka, 2009b) that valence about both the cue and the outcome might be integrated within the same neurons. However, our data suggest that this is only true for a subpopulation of DA neurons. These neurons not only decoded reward and shock omission as events of opposite valence, but also exhibited phasic activity patterns both at time of cue onset and during outcome omission that are in line

with a valence-based prediction error theory. These PE neurons were phasically excited to the reward predicting cue and shock omission, while being phasically inhibited by the shock predicting cue and reward omission. The second population of DA neurons exhibited similar modulation to outcome omissions and the two opposing cues that was extended beyond classic phasic activity. In addition, cue type was decoded very well from outcome omission firing by these neurons throughout the whole cue period, suggesting this population differentiated between cues without valence information. This indicates separate populations of DA neurons within the VTA serving different roles.

VTA DA neurons do not represent opposing appetitive and aversive systems in the brain. We did observe a small group of DA neurons that could represent an aversive system in the brain by signaling aversion through excitation and reward through inhibition. However, we found evidence against the hypothesis that VTA DA populations represent appetitive and aversive systems. The subpopulation of DA neurons that represented valence information through their activity patterns, showed very clear firing rates in line with a positive PE signal to cues predicting reward as well as shock omission and at the same time a negative PE signal to cues predicting aversive events and omission of reward. The population that did not contain valence information on the other hand exhibited large variability in firing rates to reward and shock predicting cues. This suggests that the small number of neurons that could have represented an aversive system fall into a population of neurons not representing valence at all. While this suggests that DA neurons in the VTA do not represent opposite motivational systems in the brain, it does not disprove the theory overall. Instead, appetitive and aversive systems could be represented by different structures in the brain, or different subregions within the VTA. As Lammel et. al. suggested (2011, 2014), medial and lateral areas within the VTA might contain subsets of DA neurons with differing response profiles and varying projection targets. While the recording location of DA neurons from the current

study has been confirmed to be within the VTA, the exact location of each recorded neuron along the medio-lateral axis is not known. We observed two distinct subpopulations of neurons of differing behavioral relevance. However, this does not negate the possibility of further distinct DA populations within the VTA. Future studies should further investigate firing patterns to reward and aversion predicting cues based on their projection targets and recording locations within the region.

Based on current data, the firing patterns of the valence-based PE population of DA neurons observed in this study could indicate VTA DA to represent an appetitive system. The first candidate one might think of is the amygdala, more specifically the basolateral amygdala (BLA), as decades of research have found this brain region necessary for fear learning (LeDoux et al., 1990; Tang et al., 2020; Weiskrantz, 1956; Yau et al., 2021). Some studies have found neurons in the BLA to show activity patterns in line with an aversive PE signal (Yau et al., 2021) and found dopaminergic input from the VTA modulating this signal (Tang et al., 2020), suggesting opposing integration of the two signals. Another possible candidate representing an aversive system could be the lateral habenula (LHb). Studies have found that LHb neurons exhibit firing patters in line with an aversive PE (M. Matsumoto & Hikosaka, 2009a; Stamatakis & Stuber, 2012). Additionally, LHb and VTA are integrated (Jhou, Geisler, et al., 2009) in that electrical stimulation of LHb neurons elicited inhibition of VTA DA neurons (M. Matsumoto & Hikosaka, 2007) and stimulation of LHb connections with VTA GABAergic interneurons produce conditioned place aversion (Lammel et al., 2012). A third option would lie within the VTA. GABA interneurons in this region are known to have inhibitory connections with DA neurons (Tan et al., 2012; van Zessen et al., 2012), as well as opposing activity patterns to rewarding and aversive events and cues predicting such events (Barrot et al., 2012; Jhou, Fields, et al., 2009).

Future studies should aim to further dissect these circuits and identify whether one of the above-mentioned regions or neural populations could represent an aversive system in the brain. To align with Dickinson & Dearing's theory (1979), the neurons in question would need to have inhibitory connections with VTA DA neurons, while being inhibited by DA activity in return. Additionally, excitation of the appetitive system would have to occur slightly before inhibition of the aversive system and the other way around, as an appetitive event is thought to directly activate the appetitive system, which in turn would prevent activation of the aversive system. This would lead to a delay the length of time it would take for neural transmission of the appetitive system to reach the aversive system. The same would be true for aversive to appetitive system connections.

Salience and valence signals within DA firing. While analyzing DA firing patterns in our experiment, we noticed some neurons exhibiting a very brief but strong increase in firing rate to both reward and shock predicting cues before activity diverged. This was especially apparent in the PE population. Interestingly, we also observed that both the prediction error and prediction populations showed classification accuracy similar to a random model during the first 200ms of cue presentation. Most strikingly, on both a population and by-neuron level, cue identity was not able to be determined from cue related firing during the same 200ms. This period coincides with the period of initial excitation to both reward and shock predicting cues observed in the peristimulus time histograms. Activity patterns such as initial excitation followed by inhibition during a cue predicting an aversive event have been observed by others (Fiorillo et al., 2013; Mileykovskiy & Morales, 2011). Such more complex activity patterns could hint to different temporal resolutions and signals provided by midbrain dopamine neurons. The initial response could provide information about salience, while the secondary response could signal the valence of the cue. This would mean dopamine

neurons provide a lot more information to the downstream networks than we initially thought.

Unfortunately, the current study did not include a neutral cue, to which firing to reward and shock predictive cues could have been compared. This would have provided the opportunity to investigate whether this brief excitation that could possibly be a salience signal would also be present at the start of a neutral cue. Future studies should expand on the question whether the valence PE population of VTA DA neurons signal both salience and valence. This can be achieved by recording midbrain dopamine activity during presentation of appetitive and aversive cues of different salience that are reinforced with the same valence. Here, cues would need to be controlled for both salience and sensory modality. In addition to varying the salience of cues predicting events, the modality of especially aversive events could be represented by DA neurons. A footshock might be more salient than a mild air puff and therefore evoke a more pronounced response. This could be the reason underlying the heterogeneity in findings regarding DA responses to aversive events and should therefore be investigated.

In conclusion, we were able to provide further evidence that at least a subpopulation of midbrain DA neurons represent a valence-based prediction error signal (Eshel et al., 2016; H. Matsumoto et al., 2016; Salinas-Hernández et al., 2018) in line with the classical PE model thought to underly learning (Rescorla & Wagner, 1972). We can now appreciate that DA neurons can serve multiple functions, which could be the reason underlying contradicting findings observed by studies investigating DA activity during aversion. Additionally, our data suggest that VTA DA neurons only represent an appetitive system but does not disprove a hypothesis of opposing appetitive and aversive systems. More research is required to understand how exactly aversion is represented in the brain and the impact each part of the

network has on the whole. Only then can we hope to truly understand what underlies drug addiction, anxiety and post traumatic stress disorders and how to prevent and treat them.

References

Barrot, M., Sesack, S. R., Georges, F., Pistis, M., Hong, S., & Jhou, T. C. (2012). Braking Dopamine Systems: A New GABA Master Structure for Mesolimbic and Nigrostriatal Functions. *Journal of Neuroscience*, *32*(41), 14094–14101. https://doi.org/10.1523/JNEUROSCI.3370-12.2012

Bayer, H. M., Lau, B., & Glimcher, P. W. (2007). Statistics of midbrain dopamine neuron spike trains in the awake primate. *Journal of Neurophysiology*, *98*(3), 1428–1439. https://doi.org/10.1152/jn.01140.2006

Bouton, M. E., & Bolles, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning and Motivation*, *10*(4), 445–466. https://doi.org/10.1016/0023-9690(79)90057-2

Brischoux, F., Chakraborty, S., Brierley, D. I., & Ungless, M. A. (2009). Phasic excitation of dopamine neurons in ventral VTA by noxious stimuli. *Proceedings of the National Academy of Sciences*, *106*(12), 4894–4899. https://doi.org/10.1073/pnas.0811507106

Chang, C. Y., Esber, G. R., Marrero-Garcia, Y., Yau, H.-J., Bonci, A., & Schoenbaum, G. (2016). Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. *Nature Neuroscience*, *19*(1), 111–116. https://doi.org/10.1038/nn.4191

Cheer, J. F., Aragona, B. J., Heien, M. L. A. V., Seipel, A. T., Carelli, R. M., & Wightman, R. M. (2007). Coordinated Accumbal Dopamine Release and Neural Activity Drive Goal-Directed Behavior. *Neuron*, *54*(2), 237–244. https://doi.org/10.1016/j.neuron.2007.03.021

Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B., & Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature*, *482*(7383), 85–88. https://doi.org/10.1038/nature10754

Dickinson, A. (1977). Appetitive–Aversive Interactions: Superconditioning of Fear by an Appetitive CS. *Quarterly Journal of Experimental Psychology*, *29*(1), 71–83. https://doi.org/10.1080/00335557743000044

Dickinson, A., & Dearing, M. F. (1979). Appetitive-aversive interactions and inhibitory processes. In *A. A. Dickinson & A. Boakes (Eds.) Mechanisms of learning and motivation: A memorial volume to Jerzy Konorski* (pp. 203–231). Totawa, NJ: Erlbaum.

Eshel, N., Tian, J., Bukwich, M., & Uchida, N. (2016). Dopamine neurons share common response function for reward prediction error. *Nature Neuroscience*, *19*(3), 479–486. https://doi.org/10.1038/nn.4239

Fiorillo, C. D. (2013). Two Dimensions of Value: Dopamine Neurons Represent Reward But Not Aversiveness. *Science*, *341*(6145), 546–549. https://doi.org/10.1126/science.1238699

Fiorillo, C. D., Newsome, W. T., & Schultz, W. (2008). The temporal precision of reward prediction in dopamine neurons. *Nature Neuroscience*, *11*(8), 966–973. https://doi.org/10.1038/nn.2159

Fiorillo, C. D., Song, M. R., & Yun, S. R. (2013). Multiphasic Temporal Dynamics in Responses of Midbrain Dopamine Neurons to Appetitive and Aversive Stimuli. *Journal of Neuroscience*, *33*(11), 4710–4725. https://doi.org/10.1523/JNEUROSCI.3883-12.2013

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons. *Science*, *299*(5614), 1898–1902. https://doi.org/10.1126/science.1077349

Guarraci, F. A., & Kapp, B. S. (1999). An electrophysiological characterization of ventral tegmental area dopaminergic neurons during differential pavlovian fear conditioning in the awake rabbit. *Behavioural Brain Research*, *99*(2), 169–179. https://doi.org/10.1016/s0166-4328(98)00102-8

Jhou, T. C., Fields, H. L., Baxter, M. G., Saper, C. B., & Holland, P. C. (2009). The Rostromedial Tegmental Nucleus (RMTg), a GABAergic Afferent to Midbrain Dopamine Neurons, Encodes Aversive Stimuli and Inhibits Motor Responses. *Neuron*, *61*(5), 786–800. https://doi.org/10.1016/j.neuron.2009.02.001

Jhou, T. C., Geisler, S., Marinelli, M., Degarmo, B. A., & Zahm, D. S. (2009). The mesopontine rostromedial tegmental nucleus: A structure targeted by the lateral habenula that projects to the ventral tegmental area of Tsai and substantia nigra compacta. *Journal of Comparative Neurology*, *513*(6), 566–596. https://doi.org/10.1002/cne.21891

Joshua, M., Adler, A., Mitelman, R., Vaadia, E., & Bergman, H. (2008). Midbrain Dopaminergic Neurons and Striatal Cholinergic Interneurons Encode the Difference between Reward and Aversive Events at Different Epochs of Probabilistic Classical Conditioning Trials. *Journal of Neuroscience*, *28*(45), 11673–11684. https://doi.org/10.1523/JNEUROSCI.3839-08.2008

Kamin, L. J. (1968). *"Attention-like" processes in classical conditioning.*

Kamin, L. J. (1969). Selective association and conditioning. In *Fundamental issues in associative learning* (pp. 42–64). Dalhousie University Press.

Konorski, J., & Szwejkowska, G. (1956). Reciprocal transformations of heterogeneous conditioned reflexes. *Acta Biologiae Experimentalis*, *17*(1).

Lammel, S., Ion, D. I., Roeper, J., & Malenka, R. C. (2011). Projection-Specific Modulation of Dopamine Neuron Synapses by Aversive and Rewarding Stimuli. *Neuron*, *70*(5), 855–862. https://doi.org/10.1016/j.neuron.2011.03.025

Lammel, S., Lim, B. K., & Malenka, R. C. (2014). Reward and aversion in a heterogeneous midbrain dopamine system. *Neuropharmacology*, *76*(0 0). https://doi.org/10.1016/j.neuropharm.2013.03.019

Lammel, S., Lim, B. K., Ran, C., Huang, K. W., Betley, M. J., Tye, K. M., Deisseroth, K., & Malenka, R. C. (2012). Input-specific control of reward and aversion in the ventral tegmental area. *Nature*, *491*(7423), 212–217. https://doi.org/10.1038/nature11527

LeDoux, J. E., Cicchetti, P., Xagoraris, A., & Romanski, L. M. (1990). The lateral amygdaloid nucleus: Sensory interface of the amygdala in fear conditioning. *Journal of Neuroscience*, *10*(4), 1062–1069. https://doi.org/10.1523/JNEUROSCI.10-04-01062.1990

Luo, R., Uematsu, A., Weitemier, A., Aquili, L., Koivumaa, J., McHugh, T. J., & Johansen, J. P. (2018). A dopaminergic switch for fear to safety transitions. *Nature Communications*, *9*(1), 2483. https://doi.org/10.1038/s41467-018-04784-7

Maes, E. J. P., Sharpe, M. J., Usypchuk, A. A., Lozzi, M., Chang, C. Y., Gardner, M. P. H., Schoenbaum, G., & Iordanova, M. D. (2020). Causal evidence supporting the proposal that dopamine transients function as temporal difference prediction errors. *Nature Neuroscience*, *23*(2), 176–178. https://doi.org/10.1038/s41593-019-0574-1

Mannella, F., Gurney, K., & Baldassarre, G. (2013). The nucleus accumbens as a nexus between values and goals in goal-directed behavior: A review and a new hypothesis. *Frontiers in Behavioral Neuroscience*, *7*, 135. https://doi.org/10.3389/fnbeh.2013.00135

Matsumoto, H., Tian, J., Uchida, N., & Watabe-Uchida, M. (2016). Midbrain dopamine neurons signal aversion in a reward-context-dependent manner. *ELife*, *5*. https://doi.org/10.7554/eLife.17328

Matsumoto, M., & Hikosaka, O. (2007). Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature*, *447*(7148), 1111–1115. https://doi.org/10.1038/nature05860

Matsumoto, M., & Hikosaka, O. (2009a). Representation of negative motivational value in the primate lateral habenula. *Nature Neuroscience*, *12*(1), 77–84. https://doi.org/10.1038/nn.2233

Matsumoto, M., & Hikosaka, O. (2009b). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, *459*(7248), 837–841. https://doi.org/10.1038/nature08028

Mileykovskiy, B., & Morales, M. (2011). Duration of Inhibition of Ventral Tegmental Area Dopamine Neurons Encodes a Level of Conditioned Fear. *Journal of Neuroscience*, *31*(20), 7471–7476. https://doi.org/10.1523/JNEUROSCI.5731-10.2011

Mirenowicz, J., & Schultz, W. (1994). Importance of unpredictability for reward responses in primate dopamine neurons. *Journal of Neurophysiology*, *72*(2), 1024–1027. https://doi.org/10.1152/jn.1994.72.2.1024

Mirenowicz, J., & Schultz, W. (1996). Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature*, *379*(6564), 449–451. https://doi.org/10.1038/379449a0

Nasser, H. M., & McNally, G. P. (2012). Appetitive–aversive interactions in Pavlovian fear conditioning. *Behavioral Neuroscience*, *126*(3), 404–422. https://doi.org/10.1037/a0028341

Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology*, *47*(6), 419–427. https://doi.org/10.1037/h0058775

Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., & D, H. K. (2016). Kilosort: Realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *BioRxiv*, 061481. https://doi.org/10.1101/061481

Pan, W.-X., Schmidt, R., Wickens, J. R., & Hyland, B. I. (2005). Dopamine Cells Respond to Predicted Events during Classical Conditioning: Evidence for Eligibility Traces in the Reward-Learning Network. *Journal of Neuroscience*, *25*(26), 6235–6242. https://doi.org/10.1523/JNEUROSCI.1478-05.2005

Paxinos, G., & Watson, C. (2006). *The Rat Brain in Stereotaxic Coordinates: Hard Cover Edition*. Elsevier.

Rescorla, R. A. (1981). Within-signal learning in autoshaping. *Animal Learning & Behavior*, *9*(2), 245–252. https://doi.org/10.3758/BF03197827

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement* (pp. 64–69). Appleton-Century-Crofts, New York.

Robbins, T. W., & Everitt, B. J. (1996). Neurobehavioural mechanisms of reward and motivation. *Current Opinion in Neurobiology*, *6*(2), 228–236. https://doi.org/10.1016/S0959-4388(96)80077-8

Roesch, M. R., Calu, D. J., & Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature Neuroscience*, *10*(12), 1615–1624. https://doi.org/10.1038/nn2013

Rossant, C., Kadir, S. N., Goodman, D. F. M., Schulman, J., Hunter, M. L. D., Saleem, A. B., Grosmark, A., Belluscio, M., Denfield, G. H., Ecker, A. S., Tolias, A. S., Solomon, S., Buzsáki, G., Carandini, M., & Harris, K. D. (2016). Spike sorting for large, dense electrode arrays. *Nature Neuroscience*, *19*(4), 634–641. https://doi.org/10.1038/nn.4268

Salinas-Hernández, X. I., Vogel, P., Betz, S., Kalisch, R., Sigurdsson, T., & Duvarci, S. (2018). Dopamine neurons drive fear extinction learning by signaling the omission of expected aversive outcomes. *ELife*, *7*, e38818. https://doi.org/10.7554/eLife.38818

Scavio, M. J. (1974). Classical-classical transfer: Effects of prior aversive conditions upon appetitive conditioning in rabbits (Oryctolagus cuniculus). *Journal of Comparative and Physiological Psychology*, *86*(1), 107–115. https://doi.org/10.1037/h0035966

Schultz, W. (2007). Multiple Dopamine Functions at Different Time Courses. *Annual Review of Neuroscience*, *30*(1), 259–288. https://doi.org/10.1146/annurev.neuro.28.061604.135722

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Siegle, J. H., López, A. C., Patel, Y. A., Abramov, K., Ohayon, S., & Voigts, J. (2017). Open Ephys: An open-source, plugin-based platform for multichannel electrophysiology. *Journal of Neural Engineering*, *14*(4), 045003. https://doi.org/10.1088/1741-2552/aa5eea

Sorg, B. A., & Kalivas, P. W. (1991). Effects of cocaine and footshock stress on extracellular dopamine levels in the ventral striatum. *Brain Research*, *559*(1), 29–36. https://doi.org/10.1016/0006-8993(91)90283-2

Stamatakis, A. M., & Stuber, G. D. (2012). Activation of lateral habenula inputs to the ventral midbrain promotes behavioral avoidance. *Nature Neuroscience*, *15*(8), 1105–1107. https://doi.org/10.1038/nn.3145

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., & Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience*, *16*(7), 966–973. https://doi.org/10.1038/nn.3413

Sutton, R. S. (1988). *Learning to predict by the methods of temporal differences*. *3*(1), 9–44.

Takahashi, Y. K., Batchelor, H. M., Liu, B., Khanna, A., Morales, M., & Schoenbaum, G. (2017). Dopamine Neurons Respond to Errors in the Prediction of Sensory Features of Expected Rewards. *Neuron*, *95*(6), 1395-1405.e3. https://doi.org/10.1016/j.neuron.2017.08.025

Takahashi, Y. K., Roesch, M. R., Wilson, R. C., Toreson, K., O'Donnell, P., Niv, Y., & Schoenbaum, G. (2011). Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. *Nature Neuroscience*, *14*(12), 1590–1597. https://doi.org/10.1038/nn.2957

Tan, K. R., Yvon, C., Turiault, M., Mirzabekov, J. J., Doehner, J., Labouèbe, G., Deisseroth, K., Tye, K. M., & Lüscher, C. (2012). GABA neurons of the VTA drive conditioned place aversion. *Neuron*, *73*(6), 1173–1183. https://doi.org/10.1016/j.neuron.2012.02.015

Tang, W., Kochubey, O., Kintscher, M., & Schneggenburger, R. (2020). A VTA to Basal Amygdala Dopamine Projection Contributes to Signal Salient Somatosensory Events during Fear Learning. *Journal of Neuroscience*, *40*(20), 3969–3980.

Tobler, P. N., Fiorillo, C. D., & Schultz, W. (2005). Adaptive Coding of Reward Value by Dopamine Neurons. *Science*. https://doi.org/10.1126/science.1105370

Ungless, M. A., Magill, P. J., & Bolam, J. P. (2004). Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science (New York, N.Y.)*, *303*(5666), 2040–2042. https://doi.org/10.1126/science.1093360

van Zessen, R., Phillips, J. L., Budygin, E. A., & Stuber, G. D. (2012). Activation of VTA GABA neurons disrupts reward consumption. *Neuron*, *73*(6), 1184–1194. https://doi.org/10.1016/j.neuron.2012.02.016

Vander Weele, C. M., Siciliano, C. A., Matthews, G. A., Namburi, P., Izadmehr, E. M., Espinel, I. C., Nieh, E. H., Schut, E. H. S., Padilla-Coreano, N., Burgos-Robles, A., Chang, C.-J., Kimchi, E. Y., Beyeler, A., Wichmann, R., Wildes, C. P., & Tye, K. M. (2018). Dopamine enhances signal-to-noise ratio in cortical-brainstem encoding of aversive stimuli. *Nature*, *563*(7731), 397–401. https://doi.org/10.1038/s41586-018-0682-1

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, *412*(6842), 43–48. https://doi.org/10.1038/35083500

Wang, D. V., & Tsien, J. Z. (2011). Convergent Processing of Both Positive and Negative Motivational Signals by the VTA Dopamine Neuronal Populations. *PLOS ONE*, *6*(2), e17047. https://doi.org/10.1371/journal.pone.0017047

Weiskrantz, L. (1956). Behavioral changes associated with ablation of the amygdaloid complex in monkeys. *Journal of Comparative and Physiological Psychology*, *49*(4), 381–391. https://doi.org/10.1037/h0088009

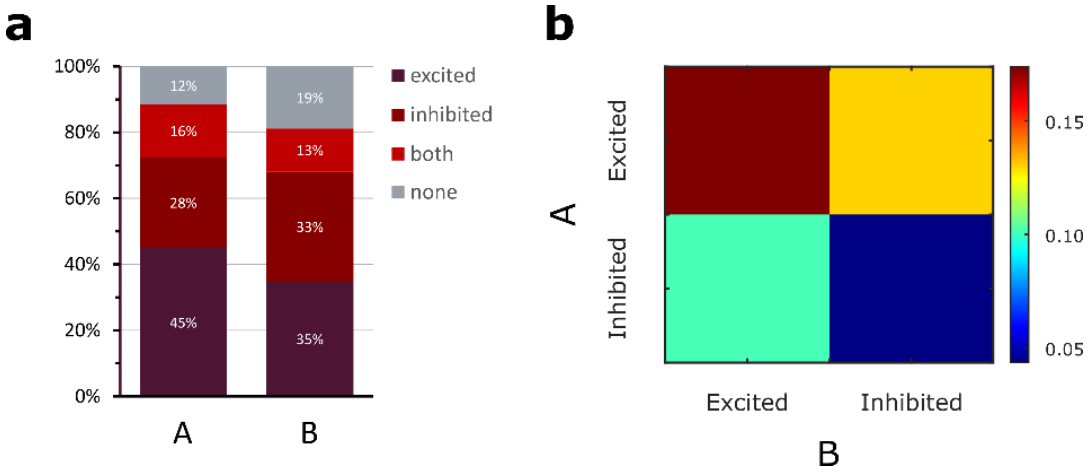
Yau, J. O.-Y., Chaichim, C., Power, J. M., & McNally, G. P. (2021). The roles of basolateral amygdala parvalbumin neurons in fear learning. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.2461-20.2021

Young, A. M. J., Joseph, M. H., & Gray, J. A. (1993). Latent inhibition of conditioned dopamine release in rat nucleus accumbens. *Neuroscience*, *54*(1), 5–9. https://doi.org/10.1016/0306-4522(93)90378-S

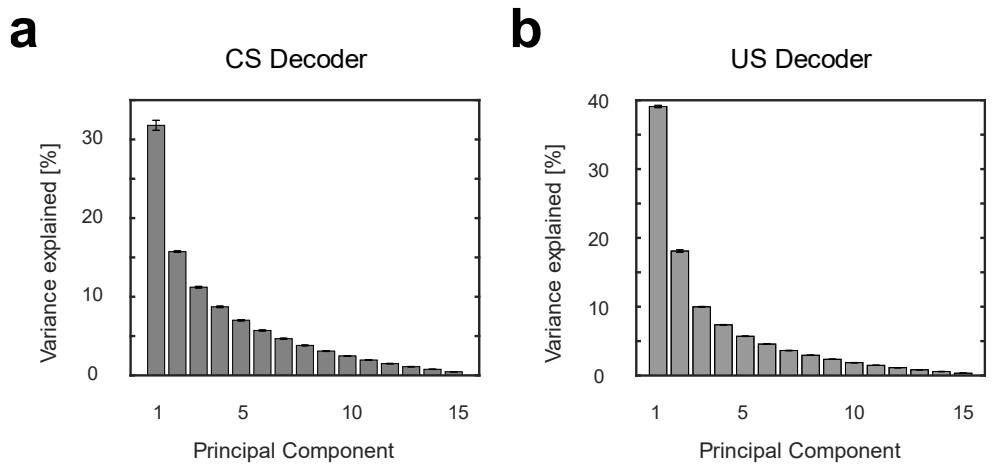
Yun, I. A. (2004). The Ventral Tegmental Area Is Required for the Behavioral and Nucleus Accumbens Neuronal Firing Responses to Incentive Cues. *Journal of Neuroscience*, *24*(12), 2923–2933. https://doi.org/10.1523/JNEUROSCI.5282-03.2004

Zweifel, L. S., Fadok, J. P., Argilli, E., Garelick, M. G., Jones, G. L., Dickerson, T. M. K., Allen, J. M., Mizumori, S. J. Y., Bonci, A., & Palmiter, R. D. (2011). Activation of dopamine neurons is critical for aversive conditioning and prevention of generalized anxiety. *Nature Neuroscience*, *14*(5), 620–626. https://doi.org/10.1038/nn.2808

Supplemental Figures



Supplemental Figure 1: DA neurons form subpopulations of activity patterns to reward and shock predicting stimuli. **a** Pro-portion of DA neurons exhibiting excitation, inhibition, both or no modulation to stimuli A and B. $X^2 = 2.66$, $p = .457$. **b** Proportion of neurons exhibiting one activity pattern to reward predicting stimulus A that exhibits the same of different activity pattern to shock predicting stimulus B. Fisher's exact test: *Odds Ratio* = 0.57, 95% CI [0.115, 2.845], $p = .697$.



Supplemental Figure 2: Variance explained by each principal component. **a** Average variance explained across principal components when PCA is conducted on FRs of CS epochs. Bars represent mean ± SEM. **b** Average variance explained by principal components when the PCA is conducted on FR during US omission. Bars presented as mean ± SEM.