Progress in Neurobiology 197 (2021) 101901

Contents lists available at ScienceDirect

Progress in Neurobiology

journal homepage: www.elsevier.com/locate/pneurobio

Original Research Article

Trial-by-trial dynamics of reward prediction error-associated signals during extinction learning and renewal

Julian Packheiser [a], José R. Donoso [b], Sen Cheng [b], Onur Güntürkün [a,1], Roland Pusch [a,1,*]

[a] Department of Biopsychology, Faculty of Psychology, Ruhr University Bochum, Universitätsstraße 150, D-44780 Bochum, Germany
[b] Institute for Neural Computation, Ruhr University Bochum, Universitätsstraße 150, D-44780 Bochum, Germany

ARTICLE INFO

Keywords:
Reward prediction error
Extinction learning
Renewal
Trial-by-trial learning
Electrophysiology

ABSTRACT

Reward prediction errors (RPEs) have been suggested to drive associative learning processes, but their precise temporal dynamics at the single-neuron level remain elusive. Here, we studied the neural correlates of RPEs, focusing on their trial-by-trial dynamics during an operant extinction learning paradigm. Within a single behavioral session, pigeons went through acquisition, extinction and renewal - the context-dependent response recovery after extinction. We recorded single units from the avian prefrontal cortex analogue, the nidopallium caudolaterale (NCL) and found that the omission of reward during extinction led to a peak of population activity that moved backwards in time as trials progressed. The chronological order of these signal changes during the progress of learning was indicative of temporal shifts of RPE signals that started during reward omission and then moved backwards to the presentation of the conditioned stimulus. Switches from operant choices to avoidance behavior (and vice versa) coincided with changes in population activity during the animals' decision-making. On the single unit level, we found more diverse patterns where some neurons' activity correlated with RPE signals whereas others correlated with the absolute value during the outcome period. Finally, we demonstrated that mere sensory contextual changes during the renewal test were sufficient to elicit signals likely associated with RPEs. Thus, RPEs are truly expectancy-driven since they can be elicited by changes in reward expectation, without an actual change in the quality or quantity of reward.

1. Introduction

Animals are able to learn which cues predict reward. This ability is the defining element of associative learning theory. The more surprising the occurrence of reward after a conditioned stimulus is, the faster the progress of learning (Rescorla and Wagner, 1972). Once the reward is fully expected after the presentation of the conditioned stimulus, learning comes to a halt. Thus, associative learning is driven by the extent of the mismatch between prediction and outcome which is known as reward prediction error (RPE). Conversely, established associations can break apart when a predicted reward is omitted after presentation of the conditioned stimulus, a process known as extinction learning (Orsini and Maren, 2012). For all this to happen, a neural signal has to be generated that propagates the mismatch between expectation and outcome to those brain areas that generate predictions about events and own action outcomes.

The signaling of RPEs has been strongly associated with the activity of dopaminergic midbrain neurons in the ventral tegmental area (VTA) and substantia nigra pars compacta (Eshel et al., 2016; Mirenowicz and Schultz, 1994; Schultz, 2016a,b; Schultz and Dickinson, 2000). These neurons demonstrate an increase in firing rate whenever unexpected reward is presented whereas they decrease their spike rate when expected reward is omitted (Schultz et al., 1997). If an animal is confronted with a completely unexpected reward, a peak of dopaminergic activity ensues. Once reward becomes more and more expected, dopamine neuron activity decreases substantially (Cohen et al., 2012; Hollerman and Schultz, 1998; Kobayashi and Schultz, 2008). Furthermore, temporally precise activations of dopaminergic VTA-signals modulate the learning rate of extinction (Steinberg et al., 2013). Thus, RPE signals are causally linked to changes in behavior and are not simply of correlative nature. Finally, RPE signals in dopamine neurons undergo a temporal shift during the incremental change in associative strength between a predictive stimulus and the associated outcome (Schultz, 2016a,b). As learning progresses, the RPE signal shifts backward from

* Corresponding author.
E-mail address: roland.pusch@rub.de (R. Pusch).
[1] These authors contributed equally to the manuscript.

https://doi.org/10.1016/j.pneurobio.2020.101901
Received 15 April 2020; Received in revised form 6 July 2020; Accepted 18 August 2020
Available online 23 August 2020

the time of reward to the time of presentation of the conditioned stimulus (Schultz, 2007). These temporal shifts are predicted by temporal difference (TD) models of learning (Sutton and Barto, 1998) and could represent the neural basis for predictive coding in the brain.

The neural correlates of RPE signaling are not limited to subcortical structures such as the VTA, but have also been found in the prefrontal cortex (Asaad and Eskandar, 2011; Oya et al., 2005) since dopamine neurons innervate the ventromedial and orbitofrontal prefrontal cortex (PFC, Slopsema et al., 1982). These areas are crucially involved in the coding for value-based decision-making and underlie economical and goal-directed behavior (Miller and Cohen, 2001). Thus, RPE signals in the PFC possibly provide feedback about past decisions and thus modify future behavior and value-based action selections (Rangel et al., 2008).

Insights about the unfolding of RPE signals on a trial-by-trial level have only recently begun to emerge. Studies in the VTA and the PFC have usually investigated this process coarsely by comparing early and late stages during learning or comparing trial blocks (Enomoto et al., 2011; Salinas-Hernández et al., 2018). However, since these RPE signals should update with every match or mismatch of the prediction, a meticulous analysis is required to gain insights into the dynamics of error signals across behavior. To capture this dynamic process, recent studies investigated how RPE signals in dopamine neurons develop with each consecutive trial during the acquisition (Coddington and Dudman, 2018; Menegas et al., 2017) and also during the subsequent extinction of a CS-US association (Pan et al., 2013). These studies found evidence that neural representations of dopamine neurons indeed update in a trial-by-trial fashion, demonstrating that granular analysis methods are tailored to track neural activity changes with high temporal precision and relate neuronal signaling to the learning process.

Since these studies used classical conditioning paradigms, the relation of RPE signal dynamics to operant behavior remains less clear. While some studies investigated RPE signals using operant conditioning tasks (Bayer and Glimcher, 2005), they did not investigate how RPE signals relate to the actual choice-behavior during the time course of learning. However, this information is paramount to understand the adaptive mechanisms of acquisition or extinction of associative learning. Lak, Stauffer and Schultz (2016a,b) for example conducted an operant learning task whilst recording from dopamine neurons during which the animals had to choose between a higher and a lower value stimulus. They found that dopamine activity was choice-dependent with increased activity levels during the choice for the higher value stimulus, a result in line with dopamine neuron signals encoding and being causally involved in upcoming value-based decisions (Morris et al., 2006; Saddoris et al., 2015). Lak et al. (2016) furthermore used trial-by-trial analyses to identify that these choice-related signal changes were not immediately present, but developed rapidly during learning.

To gain temporally precise insight into the role of prefrontal networks that have been suggested to play an important role in the RPE signal generation network (Wang et al., 2018), the present study investigated associative learning and RPE-associated signaling on trial-by-trial levels using an operant task in which acquisition, extinction and renewal test were conducted consecutively. The task was adapted from Packheiser et al. (2019a) who used this paradigm to behaviorally study the effects of memory consolidation on the renewal effect. Acquisition took place in context A, extinction in context B, and renewal test again in context A. Thus, our paradigm constitutes a classic ABA renewal design (Bouton, 2004). We chose this paradigm to track single neuron activity over multiple stages of learning to reveal the trial-by-trial evolution of RPE signals and the trial-by-trial dynamics of decision making of the animals. Pigeons were used since they easily learn complex associative learning paradigms over multiple stages (Starosta et al., 2014). We recorded single unit activity from the nidopallium caudolaterale (NCL), the avian analogue to the PFC (Güntürkün, 2005). Comparable to the PFC in mammals, the avian NCL is strongly innervated by dopaminergic neurons from the VTA (Kröner and Güntürkün, 1999; Wynne and Güntürkün, 1995), displays RPE signals (Lengersdorf et al., 2014a,b) and is strongly involved in decision-making (Veit and Nieder, 2013; Veit et al., 2015). For that reason, we hypothesize that NCL neurons track both the development of reward prediction error-associated signals during extinction learning and renewal as well as the neural correlates of upcoming decisions in a trial-by-trial fashion.

2. Results

Behavioral and neural data were obtained in 56 recording sessions from eight pigeons. Prior to the experimental sessions, a shaping and pre-training procedure was conducted in which the animals were habituated to the experimental procedure that entailed the learning of a discrimination between two previously unknown stimuli. Furthermore, they were pre-trained onto two familiar stimuli that were constant in each session during pre-training and that served as controls (see Methods). After reaching a performance criterion in the pre-training phase, the animals underwent surgery and were then exposed to the full experimental procedure. Here, the animals went through an appetitive operant acquisition, followed by extinction and subsequent renewal test (Fig. 1A) in each individual session. During acquisition in context A, pigeons had to learn stimulus-response (S-R) associations for two novel stimuli by trial-and-error by acquiring which stimulus had to be associated with a peck on the left or the right response key (identical to the pre-training procedure). In interspersed manners, the two familiar control stimuli were presented. After reaching the acquisition performance criterion, the extinction phase started immediately: One of the novel stimuli was randomly chosen to be extinguished and was no longer rewarded or punished irrespective of the animal's choice. The reward contingency of the other novel stimulus did not change throughout the experiment, which served as a further control stimulus for which the S-R association had only recently been learned. The extinction phase took place in context B as defined by a different ambient light color in the experimental chamber. After reaching the extinction criterion, the extinction phase ended. The context switched back to the acquisition context A to induce renewal, i.e. the return of the conditioned response. During the test for renewal, reward contingencies were identical to the extinction phase to study whether the previously extinguished conditioned response returns only due to the contextual change rather than a return of the reward.

Every trial started with a 4 s intertrial interval (ITI, Fig. 1B: acquisition, Fig. 1C: extinction trials). Then, the animal had to peck the center key to start the *stimulus presentation period* (2.5 s) during which one of the four stimuli was shown. After a confirmation peck, the *choice period* started where the animals had up to 6 s to make a left or right choice. Subsequently, the 2 s *outcome period* started, in which correct choices earned food reward while incorrect choices resulted in a 2 s time-out. We recorded the number of conditioned choices as well as the number of pecks onto the sample stimulus, a Pavlovian measure, as dependent variables. To compare behavioral data across sessions, experimental phases were divided into equally sized six blocks for analysis.

2.1. Behavioral results

We now will describe first the behavioral and then the neural results. Animal-specific information on the behavioral sessions, i.e. number of neurons recorded per animal and session as well as the average number of trials per phase, are depicted in supplementary table 1. Since we were specifically interested in the trial-by-trial dynamics at the phase transitions "acquisition → extinction" and "extinction → renewal", we will focus on these parts. Additional information on the acquisition phase can be found in the supplementary results.

2.2. Phase transition: acquisition → extinction

To investigate changes in choice behavior due to the switch to extinction, we compared the last block of acquisition to the first block of

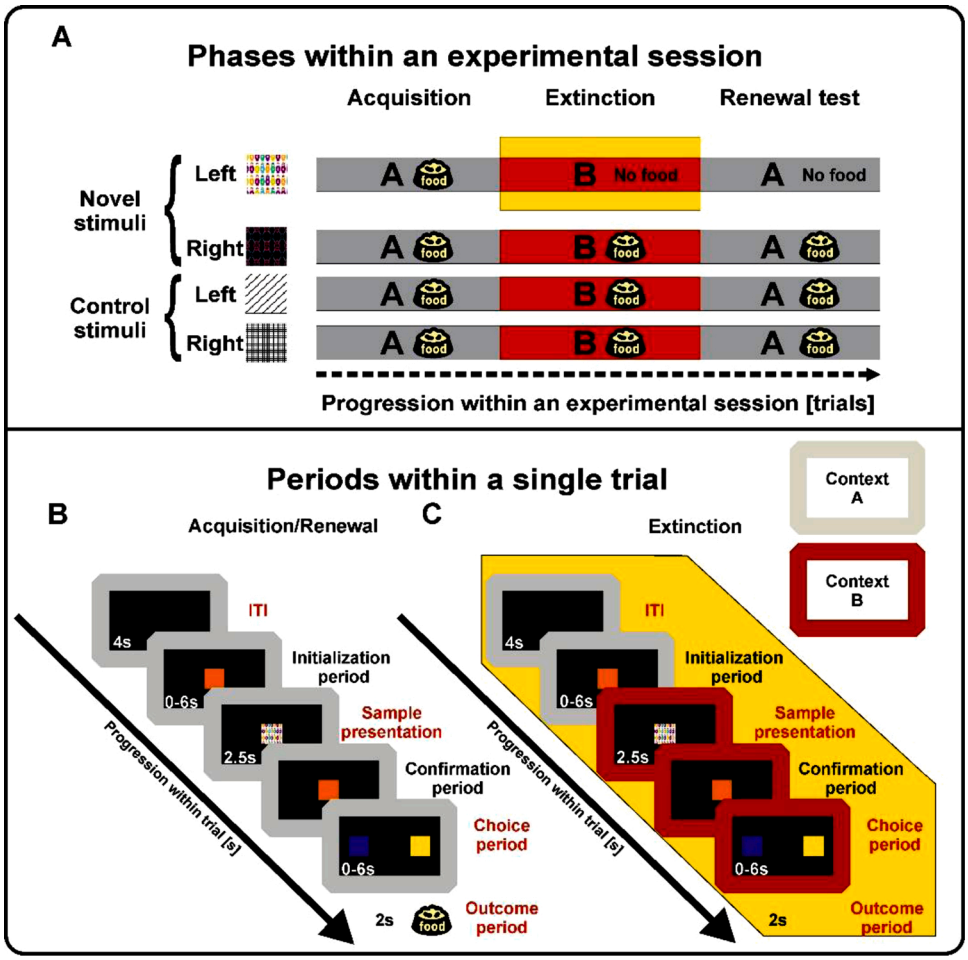

Fig. 1. Experimental design. A) Overview of the experimental procedure. During an initial acquisition in context A, the animals learned the association between left or right choices with two novel visual stimuli. After reaching learning criterion, extinction started in context B. Here, one of the novel stimuli was randomly chosen to be extinguished (highlighted in yellow). To investigate the renewal effect in a final renewal test, the acquisition context A was re-established, while pecking the extinction stimulus still did not elicit reward. B) Trial structure during acquisition and the renewal test trials. In the acquisition and renewal test phase, the trials took place under house light conditions (context A, indicated by the gray frame). After an ITI, the animals were required to initialize the trial by pecking. During the sample presentation, one of four experimental stimuli was shown on the center of the screen. A center peck confirmation triggered the illumination of the side keys. In the choice period, the animals were required to make a choice in accordance with the preceding stimulus. Finally, during a 2 s long outcome period either reward, punishment or no feedback occurred. C) Trial structure during extinction trials. A successful initialization triggered a change in the light conditions. The red rectangle at each step illustrates the change to context B. The context light remained on until the end of the trial. Note that the outcome period remained void of any feedback for 2 s regardless of the animal's decision in extinction stimulus trials (highlighted in yellow).

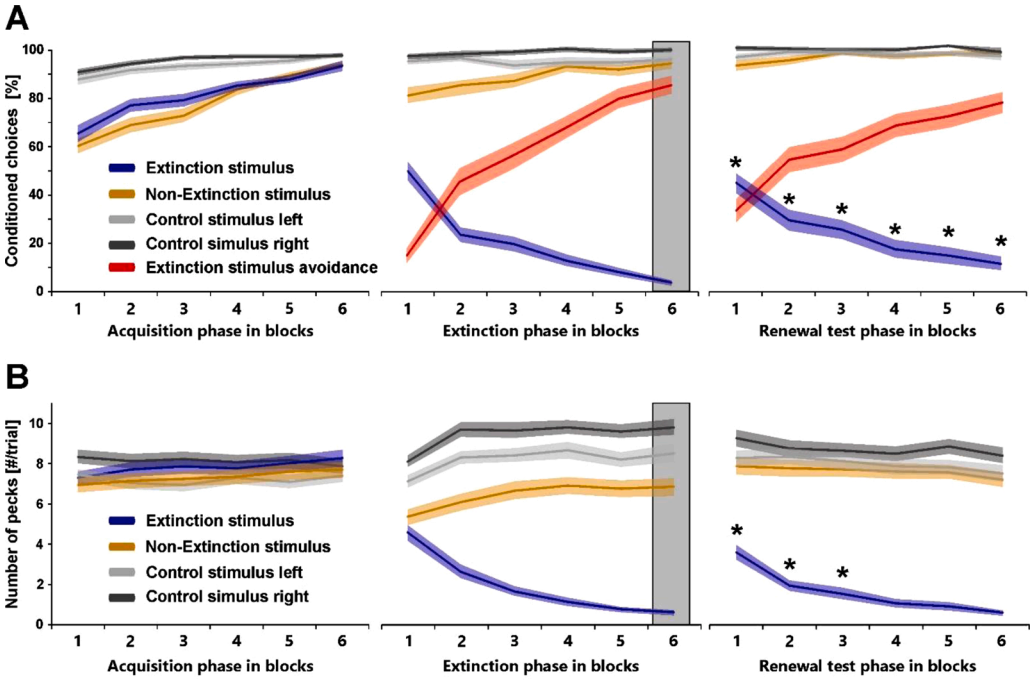

Fig. 2. Averaged behavioral results across the 56 recordings sessions. A) Conditioned choices for all four stimuli are plotted for the three stages of learning (left = acquisition, center = extinction, right = test phase). Since the length of the experimental phases was variable per session due to the behavioral criteria (cf. SI table 1), each phase was subdivided into six even blocks. During extinction and the test phase, active avoidance of the upcoming choice is plotted for the extinction stimulus only. The last block of the extinction phase is highlighted since it served as reference for the extent of the renewal effect. Blocks exhibiting a significant renewal effect in the test phase are marked by asterisks. B) As in A), but for pecking frequencies elicited by each stimulus during the sample presentation. Shaded areas represent SEM.

extinction. Here, we found a significant reduction in conditioned choices and pecking rates for the extinction stimulus (performance: $t_{(55)} = 9.41$, $p < .001$; pecks: $t_{(55)} = 9.17$, $p < .001$) and, albeit to a much smaller degree, the non-extinction stimulus (performance: $t_{(55)} = 5.20$, $p < .001$; pecks: $t_{(55)} = 2.26$, $p = .028$, Fig. 2A and B, center panels). While the response reduction to the extinction stimulus was expected, a corresponding finding for the other novel stimulus hints at an initial generalization between both novel stimuli at extinction onset.

To investigate trial-by-trial dynamics of extinction learning, we applied a sliding window analysis to all responses to the extinction stimulus (Gallistel et al., 2004). Individual trials during extinction were classified as either "*persistent behavior*", "*exploratory behavior*" or "*avoidance behavior*" (see Methods and SI Fig. 1 for details). A behavior in a trial was classified as "persistent" if at least three out of five conditioned responses were directed to the previously rewarded choice key. A behavior in a trial was classified as "avoidance" if at least three out of five responses were choice omissions. Behavior was labeled "exploratory" if it was neither classified as persistent or avoidance. Based on this classification, we calculated the median onset of these individual behaviors during the extinction phase for each behavioral session.

Animals displayed "persistent behavior" at the onset of extinction in all 56 behavioral sessions for the extinction stimulus (median first occurrence = trial 1). "Exploratory behavior" commenced after the animals encountered a few trials without reward (median first occurrence = trial 5). "Avoidance behavior" emerged after exploratory behavior did not yield any reward for the animals (median first occurrence = trial 13). The sequence of the behaviors was identical across sessions: extinction learning started with persistent behavior, followed by exploratory behavior and then avoidance behavior in 55 out of 56 sessions. Example sessions are shown in SI Fig. 1 and SI Fig. 2.

2.3. Phase transition: extinction → renewal

We used the last block of extinction (gray shaded area in Fig. 2A & B, center panel) as reference for the extent of renewal measured during the renewal test. After the switch back to context A, we indeed recorded more conditioned choices for the extinction stimulus in all six extinction renewal test blocks than in the last extinction block in context B ($p < .008$, Fig. 2A, right panel). We found similar results for pecking rates for the first three blocks ($p \leq .001$, Fig. 2B, right panel). No differences were significant for the other stimuli (all p's > 0.250).

To see if renewal behavior started abruptly, we compared the number of conditioned choices in the last extinction trial with the first trial during renewal. Indeed, only in a single out of all 56 extinction sessions, we found a conditioned response in the last trial of extinction. In contrast, a conditioned choice was observed in the first trial significantly more often in 24 out of 56 renewal sessions ($\chi^2_{(1)} = 27.24$, $p < .001$).

In short, extinction onset first induced exploratory behavior, followed by a continuous decrease of responses to the extinction stimulus. The return of conditioned responding during the renewal test was immediate and visible during the very first trial of about half of the sessions. In a next step, we investigated if avian "prefrontal" neurons displayed trial-by-trial dynamics of activity patterns corresponding to these changes in behavior.

2.4. Electrophysiological results

A total of 136 NCL neurons were recorded (SI Fig. 3 for recording sites) to investigate the temporal dynamics of neural activity during extinction learning. We conducted a sliding window comparison of firing rate changes at the border of phase transitions (see Methods for details). For the neural data, we used a 10-trial wide sliding window since neuronal firing rates are more variable and require pooling across a larger window for reliable assessment. In a first step, raw firing rates for each cell were obtained for four distinct periods within each trial and separately for all four experimental stimuli: ITI; sample presentation; choice period; outcome period (Fig. 1 B&C, highlighted in red). To compare neural responses across the population, raw firing rate changes were z-transformed. All calculations are based on these normalized firing rates.

Then, we individually quantified changes in the firing rate of all 136 recorded cells at the transition "acquisition → extinction". Since neural activity is variable even in the absence of overt changes in the environment, a "baseline" condition was required to measure if the activity changes in the experimental periods were meaningful. We therefore extracted the baseline change for each cell from a period in which we would not expect systematic changes related to learning, i.e. during the ITI. Here, we only expected stochastic fluctuations at the phase transition (see Fig. 3A). Our baseline condition was calculated as follows:

$$Baseline\ \Delta_i^k = |\overline{C}_i^k - \overline{R}_i|$$

Here, $\overline{C}$ represents the average firing rates of a comparison window (i.e. the first 10 trials in an experimental phase; see Fig. 3A). $\overline{R}$ represents the average firing rates in a fixed reference window (last 10 trials of the preceding experimental phase). For each recorded neuron i, we calculated the difference Δ_i between $\overline{C}$ and $\overline{R}$. Since firing rate changes could be positive or negative after the phase transition, Δ_i was transformed into absolute values. $|\Delta_i|$ was calculated consecutively for every succession of the comparison window (k). The index k always started at the first trial of an experimental phase and continued to increase trial-by-trial.

The identical analysis was then repeated for all cells in the three experimental periods (sample presentation, choice period, outcome period) using the same formula:

$$Exp\ \Delta_i^k = |\overline{C}_i^k - \overline{R}_i|$$

After quantification of the individual firing rate changes during baseline and each experimental period for each cell, we examined whether these changes were significant on the population level. To this end, we compared the baseline signal change during ITI ($Baseline\ \Delta_i^k$) to the activity during each experimental period ($Exp\ \Delta_i^k$) using paired t-tests. This calculation was performed for every succession of the comparison window (k). The resulting single value for each k is defined as the net activity change over time. This value indicates the strength of the population firing rate change at the phase transition for the respective experimental period (Fig. 3B 1st consecutive comparison).

Net activity changes were expressed as measures of effect size (Cohen's d) and are separately illustrated for each stimulus as color-coded graphs (a hypothetical example is given in Fig. 3B). We chose effect sizes rather than p-values since effect sizes provide direct insights into the strength of a difference rather than dividing comparisons into dichotomous significant/non-significant results. Only at least small to medium effects were regarded as meaningful (Cohen's d > 0.3). This threshold was determined by conducting a permutation test with shuffled data (see Methods and SI Fig. 4 for further information).

2.5. Phase transition: acquisition → extinction

We investigated neural signal changes of the NCL population at the transition from acquisition to extinction by analyzing 30 comparison windows to cover both the early and late extinction phase. Activity changes per individual neuron during extinction for extinction stimulus trials are presented in SI Fig. 5 (stimulus presentation), SI Fig. 6 (choice period) and SI Fig. 7 (outcome period).

3. Stimulus presentation

During the stimulus presentation period of the extinction phase, net activity changes occurred at two time points (Fig. 4A). A minor one

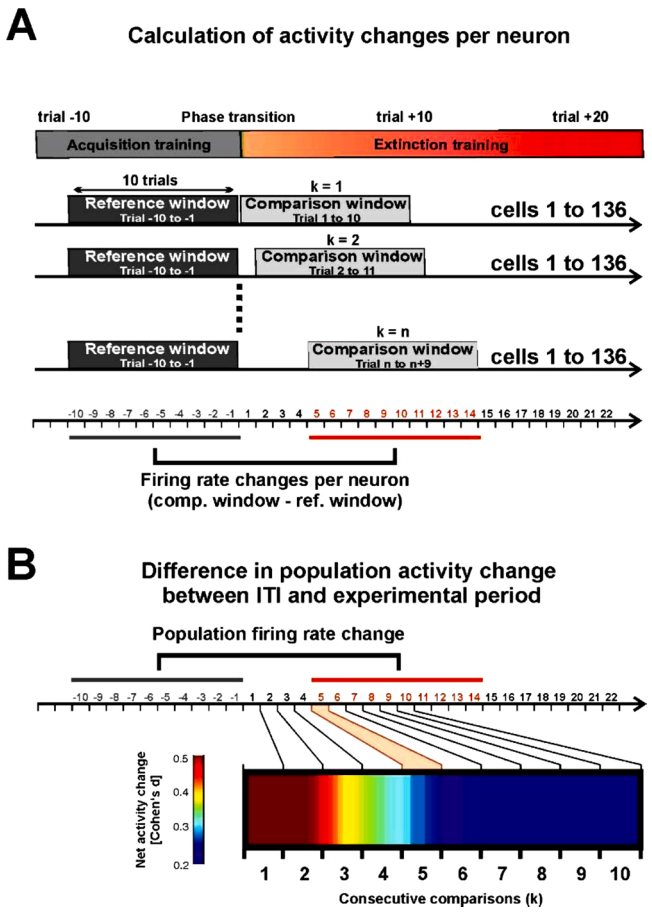

Fig. 3. Trial-by-trial sliding window analysis to quantify net activity changes across experimental phases. A) For each cell, we first quantified changes in firing rates at the transition from the end of acquisition (last 10 trials, reference window) to the beginning of extinction (first 10 trials, comparison window at k = 1) during the ITI and each experimental period (e.g. the stimulus presentation). The comparison window was then moved through the extinction phase on a trial-by-trial basis (k = 1, k = 2, … k = n). Firing rate changes per neuron were calculated by subtracting the mean firing rate during the reference window from the mean firing rate in the comparison window. B) To assess changes in the neuronal population, we compared the firing rate changes between the baseline condition and the three experimental periods for all 136 cells using paired t-tests. These comparisons were conducted for each comparison window to identify population changes in a trial-by-trial fashion. Here, we present hypothetical data to illustrate the method. In this example, effect sizes of firing rate changes were strong at the beginning of extinction, indicating a robust change in firing rate during the experimental period (e.g. the stimulus presentation) compared to the baseline condition (ITI). These changes diminished over the course of extinction. The 5th comparison window is highlighted to illustrate how the 10-trial window maps onto a discrete data point in the figure.

started at the 8th comparison window whereas a delayed population response to extinction occurred at the 20th comparison window. This change is also visible in individual neurons (Fig. 5). Consistent with our behavioral results and previous reports (de Voogd et al., 2020), net activity changes also occurred for the non-extinction stimulus, possibly due to stimulus generalization across novel stimuli. No changes were observed for the two familiar stimuli.

4. Choice period

NCL neurons encode upcoming choices prior to the animal's choice peck (Lengersdorf, Pusch et al., 2014; Starosta et al., 2014; Veit et al., 2015, 2014; Veit and Nieder, 2013). For that reason, we analyzed the choice period to investigate neural activity changes in relation to changes in decision-making. We found a delayed net activity change during the choice period of extinction learning starting from the 6th comparison window (Fig. 4B). This accords with the observation that animals started exploratory behavior about 5 trials after extinction onset. At about trial 13, our pigeons switched to avoidance behavior, although switch time was highly variable between individuals and sessions (SD avoidance onset = 12.47 trials). The net activity changes also exhibited a second peak, starting in the 21st comparison window. Possibly, this delayed onset of neural change was related to the higher inter- and intraindividual behavioral variability in avoidance behavior onset. Consistent avoidance behavior manifested only late during extinction (see SI Fig. 2 for an example session). Because of the substantial session-to-session variability in the behavior, we tested whether the changes in neural activity during the choice period were indeed associated with changes in behavior in individual sessions. To this end, we compared firing rates in the last 10 trials prior to the first 10 trials after the onset of exploratory behavior thus considering the session-to-session variability. The same approach was applied to the onset of avoidance during extinction. Choice period activity did not change significantly when behavior switched from persistent to exploratory behavior ($p > .250$), but changed significantly at the switch from exploratory to avoidance behavior ($p = .021$). A minor change occurred for the left control stimulus at the 12th comparison window. Net activity changes for the other stimuli were below threshold.

5. Outcome period

Finally, we analyzed net activity changes during the outcome period, i.e. the time in the trial when the reward was either presented (during acquisition) or omitted (during extinction). Here, we found net activity changes in the very first comparison window for the extinction stimulus (Fig. 4C). To identify if they resulted from a reward prediction error, we analyzed if they disappeared during late extinction stages when reward omission was then fully predicted by the animal. This was indeed the case from the 24th comparison window onward. Thus, these results provide first indications that NCL neurons as a population seem to encode RPE signals (see Fig. 5 for an individual neuron).

A sign change in neural activity from positive to negative RPEs is a constitutive feature of RPE signaling (Schultz, 2015). To provide further insights into whether the signal changes could have been constituted by RPEs, we thus investigated whether a sign change in the signal of individual neurons occurred from the end of acquisition to the beginning of extinction (see Methods). Since there was possibly still a residual positive RPE at the end of acquisition, the switch to a negative RPE during extinction should have elicited an inversion of signal change if NCL neurons encode RPEs. Indeed, we found that 19 individual cells changed their signal strength (Cohen's $d > 1$ equaling a large or very large effect) and changed sign at this phase transition. In a secondary analysis, we investigated whether activity changes in these neurons decreased over the course of extinction learning as would be expected if these neurons represent RPEs. The expected decline in signal change was observed in a majority (14/19 or 74 %) of these neurons.

Since not all neurons demonstrated a "fading" of activity changes during later stages of the extinction phase, we quantified how many cells demonstrated activity changes of at least Cohen's $d > 0.5$ throughout all 30 comparison windows to identify neurons that possibly carried information about the continued absence of reward. Here, 25/136 or 18 % of the neurons showed a continuous change in firing rate across the extinction phase. These neurons might thus represent the absolute value of reward (present vs. absent) rather than RPE signals.

In summary, we found significant net activity changes during the outcome period that emerged immediately upon onset of the extinction phase. These changes disappeared with ongoing learning and then moved to the stimulus presentation period on the population level. Neural activity during the choice period changed significantly around the time when behavior switched from exploratory to avoidance

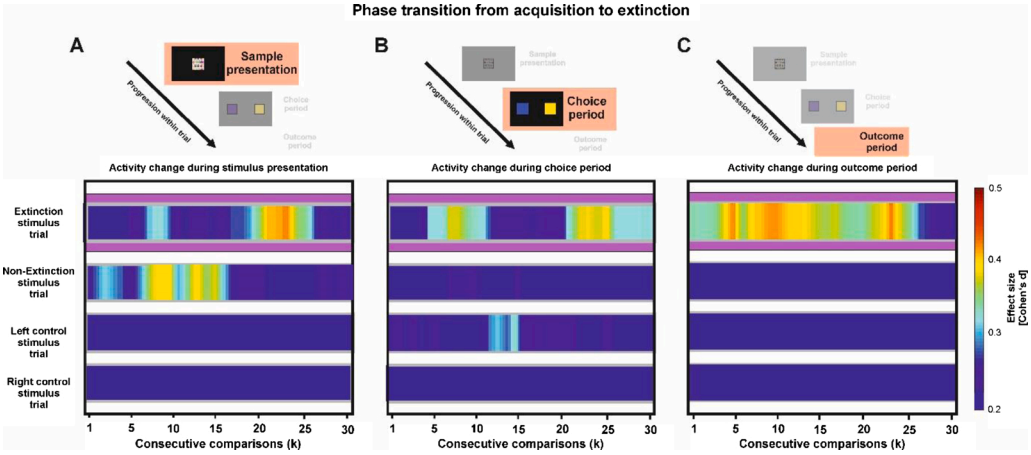

Fig. 4. Net activity changes in the NCL population emerge sequentially at the transition from acquisition to extinction. The analyzed experimental periods are highlighted on top. A) For the extinction stimulus, net activity changes during stimulus presentation started modestly in the 8th comparison window and had a second strong peak starting at the 20th comparison window. Net activity changes also occurred for the non-extinction stimulus trials (second from top). No activity changes occurred for the control stimuli. B) Net activity changes during the choice period likely reflect premotor signals of the upcoming decision. For the extinction stimulus, these signals started at the 5th comparison window and had a second peak starting from the 20th comparison window. C) For the extinction stimulus, net activity changes during the outcome period were immediately present after the extinction phase started for extinction stimulus trials. During late extinction, these changes disappeared, indicative of RPE signaling.

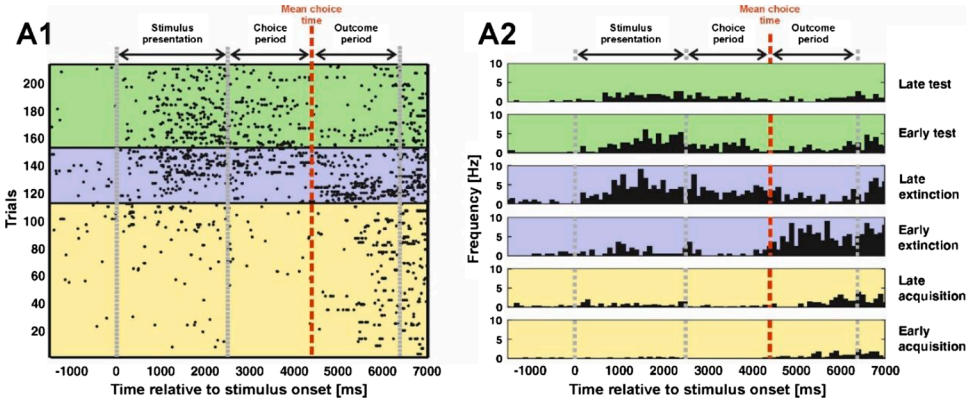

Fig. 5. Example of a single unit that exhibits sequential changes in different trial periods for the extinction stimulus. A1) Raster plot for the extinction stimulus. Time within trial is plotted on the x-axis with stimulus onset serving as point zero. Phase borders within the trial are marked by dashed vertical lines. The mean choice time is plotted separately in red to provide an estimate when choices occurred on average during the whole session. The y-axis indicates the number of trials the extinction stimulus was presented during the experimental session. Yellow, blue and green backgrounds indicate the acquisition, extinction and the renewal test phase, respectively. After extinction onset, this neuron increased its firing rate during the outcome period likely due to the unexpected omission of reward (reward prediction error). Over the course of extinction, this activity pattern moved to the stimulus presentation, indicating that the presumed RPE signal moved to the predictive cue. The raster plot for all other experimental stimuli are shown in SI figure 8. A2) PSTHs for the extinction stimulus. PSTHs always depict the first and second half of each experimental phase to illustrate dynamic changes.

behavior. Activity changes in individual neurons during the outcome period were more diverse. A subset of neurons switched sign at the phase transition and decreased their activity changes with ongoing extinction, another subset showed continuous changes in activity during the outcome period throughout extinction.

5.1. Phase transition: extinction → renewal

We next analyzed the phase transition from extinction (context B) to the renewal test phase (context A). Here, the end of extinction was used as reference window and compared to the sliding comparison window throughout the renewal test. We focused on the first 10 comparison windows as renewal was most pronounced immediately after context switch.

We could not find any meaningful changes in neural activity if all neurons were included in the analysis (see SI figure 9). The most obvious explanation for a lack of activity differences was that the extent of renewal differed between individual sessions similar to findings in humans (Lissek et al., 2016). We therefore decided to only analyze sessions with clear renewal. Using a median split for the conditioned choices in the first renewal block, 26 behavioral sessions were classified as renewal sessions (median number of conditioned choices = 45.45 %). Of the in total 136 recorded neurons, 69 cells were recorded in these renewal sessions and are included in the subsequent analyses.

6. Stimulus presentation

For the stimulus presentation period, we found a small net activity change in the first comparison window (Fig. 6A for population and SI figure 10 for individual neuron response) that increased as the renewal test progressed. Possibly, the extinction stimulus did not lose all its associative strength during extinction learning, and conditioned responses ceased because they were suppressed by a newly learned inhibitory association with the extinction context (see SI figure 11 for a model of associative strength). The ensuing re-extinction then resulted in a further loss of associative strength.

7. Choice period

Net neural activity changes in the choice period were already visible in the first comparison window, similar to the observed behavioral changes (Fig. 6B for population and SI figure 12 for individual neuron response). Fig. 7 illustrates a cell that tracks extinction and renewal during the choice period in extinction stimulus trials. The analysis of error trials illustrates that this neuron was signaling the upcoming decision during the choice period (SI figure 13).

8. Outcome period

Renewal cells yielded net activity changes during the outcome phase in the first comparison window (Fig. 6C for population and SI figure 14 for individual neuron response). Since no changes in reward contingency had occurred, this effect was likely driven by a reward expectancy violation induced by the change from extinction context B back to acquisition context A. Accordingly, this net activity change dissipated from the 5th comparison window onward. No other stimulus produced above threshold signal changes.

9. Discussion

In the current study, we investigated the dynamics of reward prediction error (RPE)-associated signals and decision-related coding in the NCL - the avian analogue of the prefrontal cortex. For this purpose, we employed an appetitive ABA extinction learning paradigm. A trial-by-trial sliding window analysis revealed that the NCL population showed altered activity levels at the immediate onset of extinction during the omission of reward. These activity peaks disappeared from the time of reward omission and re-appeared during stimulus presentation. In addition, we observed for the first time neural signatures of renewal at the population level. The renewal effect could be observed immediately after the contextual change. Here, we found swift neural signal changes for outcome, choice- and stimulus-related activity that coincided with instant changes in behavior. Our fine-grained analysis of the activity changes revealed the temporal cascade of coding events during learning. Furthermore, it indicates that RPE-associated signals can be elicited by a mere contextual change that prompts a return of reward expectancy for an extinguished stimulus.

Both during "acquisition → extinction" and "extinction → renewal" transitions, we found temporary activity changes in the outcome period during extinction. These activity changes possibly represent RPE-associated signals as they diminished over time, indicating that the NCL at the population level encodes rewards in relation to their expectation rather than as absolute values. To further elucidate whether the neuronal changes observed at the population level were associated with RPEs, we investigated if single neurons in the population exhibited a sign change from the end of acquisition to the beginning of extinction. By definition the RPE is a signed value and should change from positive RPEs during unpredicted reward delivery, to negative RPEs during the unpredicted reward omission at the onset of extinction (Schultz, 2015). Although our behavioral task did not feature the presentation of unexpected reward specifically, we assumed some positive RPE during the outcome period since the newly formed within-session association between stimuli and reward was likely not fully predicted yet. Indeed, we found a subset of neurons switching their sign between these two phases. Further, they decreased their signal change from the beginning to the end of extinction, suggesting that they could have represented RPEs. Our results thus provide additional indications that "prefrontal" regions in both birds and mammals encode signals associated with RPEs (e.g., Asaad and Eskandar, 2011; Lengersdorf, Pusch et al., 2014; Oya et al., 2005). Since RPEs are considered a learning signal (Schultz, 2016a,b), silencing prefrontal or prefrontal-analog neurons would likely lead to impairments in extinction learning. This interpretation is in full accordance with results from studies with monkeys (Walker et al., 2009) and pigeons (Lengersdorf et al., 2015). In both studies, pharmacological inactivation of the PFC and NCL during extinction learning resulted in a severe impairment of the extinction learning process.

We furthermore investigated the temporal shifts from US to CS that have been suggested by TD models of learning (Sutton and Barto, 1998) at trial resolution. We found that signal changes during extinction were initially present at reward omission but disappeared with ongoing learning at the population level. With their disappearance during the outcome period, they emerged during the time when the predictive stimulus was presented. These results at the population level are in line

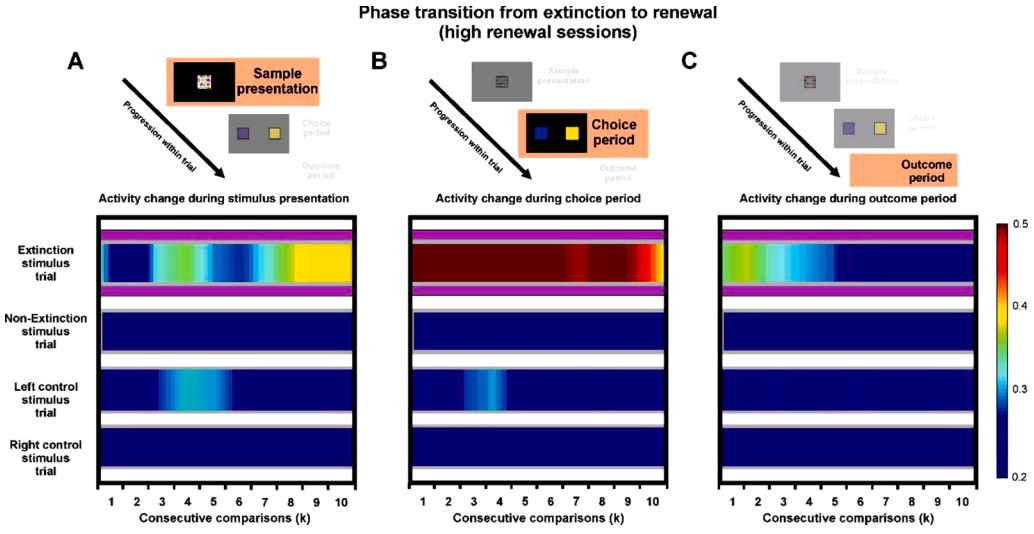

Fig. 6. Net activity changes emerge immediately in all trial periods at the transition from extinction to the renewal test phase for the extinction stimulus. A) Net activity changes during the stimulus presentation were immediate during extinction stimulus trials, but quickly faded. Net activity changes then reappeared and increased as the test phase progressed. In addition, a small effect was present for the left control stimulus. B) Net activity changes during the choice period were also immediately present for extinction stimulus trials corresponding to immediate changes in behavior. C) Net activity changes during the outcome period were immediately present for extinction stimulus trials during the renewal test even though the extinction stimulus remained unrewarded. Again, these changes disappeared with ongoing learning.

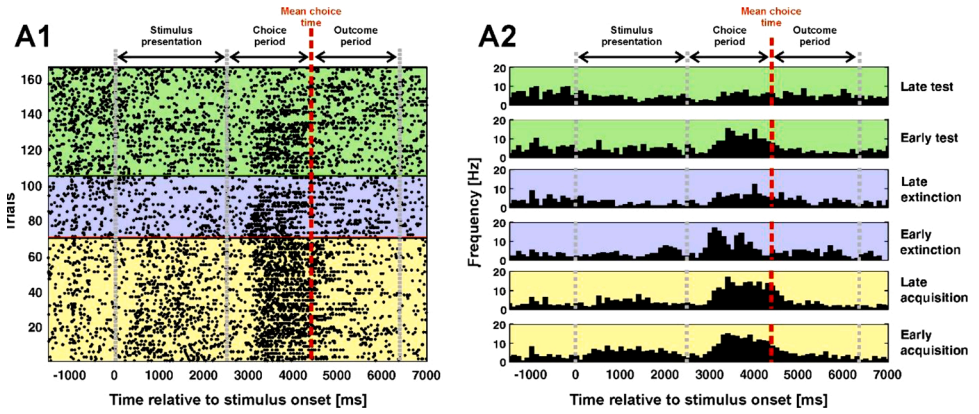

Fig. 7. Neuronal response of an example single unit during extinction stimulus trials that was preferably tuned to a left-sided choice. A1) Raster plot for the extinction stimulus that was associated with a left-sided choice in this experimental session. During extinction, the activity pattern ceased due to avoidance behavior. After returning to acquisition context A, the firing pattern from acquisition and early extinction reappeared signaling renewal. The raster plot for all other experimental stimuli are shown in SI figure 15. A2) PSTHs for the extinction stimulus.

with results from Sadacca et al. (2016) as they demonstrated a gradual shift in dopamine neuronal activities from the outcome presentation to the cue that predicted reward. Similarly, Menegas et al. (2017) observed a gradual shift in population dopamine axon activity in the ventral striatum over the course of several sessions measured via calcium imaging. In the cited study, dopamine activity only decreased rather than faded during US presentation despite being fully predicted by the CS. In contrast to our experiment, recordings of both studies were conducted over the course of several days of conditioning and thus occluded the trial-by-trial dynamics of learning. Our model organism allowed for the investigation of CS-US associability in individual sessions as pigeons can perform over 1000 trials each day (Starosta et al., 2014) enabling the investigation of multiple learning steps. Our results also suggest that changes in error signals are gradual, on the trial rather than session level, supporting models of associative learning in which changes in error signaling are gradual as well and update on a trial-by-trial basis (Gluck and Bower, 1988; Rescorla and Wagner, 1972).

While the results are in line with TD models at the population level, we observed diverse neuronal responses during the outcome period at the single unit level (see SI figure 9). 18 % of NCL neurons exhibited continuous changes in activity throughout the entire extinction phase that did not disappear and sometimes even increased over time. Interestingly, a subset of dopamine neurons exhibited comparable firing patterns indicating that they were not simply RPE coding neurons but represent variables such as an absolute/pure reward code or motivational salience (Matsumoto and Hikosaka, 2009). It seems reasonable that the information about the presence or absence of reward should be represented on the neural level in addition to the representation of the prediction as the degree of error can only be computed if the outcome itself is taken into account (for review see Watabe-Uchida et al., 2017).

One key advantage of our behavioral paradigm was that we were able to investigate the neural correlates of expectancy violation without the modulation of quantity or quality of reward during the renewal test. We discovered that NCL population activity immediately changes for stimulus, choice and outcome-related coding at the onset of renewal. Since the absolute outcome value did not change from extinction to renewal test, these signal changes cannot be attributed to a change in absolute reward. This change in population activity could be an indication of RPE-associated signaling as it is purely driven by the mismatch of the reinstated reward expectation through the context change and the continuous outcome omission in extinction stimulus trials.

While there were indications that activity changes at both the single unit as well as the population level were constituted by RPE signaling, it cannot be ruled out that these signals represented variables that are known to co-occur during expectancy violation such as novelty (Menegas et al., 2017). Since we discovered that NCL population activity immediately demonstrated signal changes for stimulus, choice and outcome-related coding at the onset of renewal, another possible explanation for the observed changes could be attributed to the change in the ambient context per se. It is however very unlikely that the mere physical color change drove these activity changes as they were identical for all experimental stimuli. As the context change only affected neural activity in extinction stimulus trials, it seems more probable that the NCL population encodes stimuli as predictors for the outcome of the organism's own behavior. The overall sensory properties of the different contexts could instead primarily be encoded in visual associative areas as previously shown (Gao et al., 2019; Lengersdorf et al., 2014a,b). It is conceivable that these structures propagate contextual information to the NCL depending on the animal's decision-making process, i.e. only when they become behaviorally relevant.

An important aim of the present study was to investigate changes in decisional coding in the prefrontal NCL. At the behavioral level, we discovered that extinction learning first induced choice variations before resulting in a stop of responding. This is similar to other findings in humans and other animals (Eckerman and Lanson, 1969; Kinloch et al., 2009; Neuringer et al., 2001; Rick et al., 2006). On the neural level, NCL signal changes during the choice period are known to strongly correlate with upcoming behavior (Lengersdorf, Pusch et al., 2014; Starosta et al., 2014; Veit et al., 2014, 2015; Veit and Nieder, 2013). Indeed, we found significant alterations in the population activity during the choice period. These were found at the transition from exploratory to avoidance behavior during extinction as well as from avoidance to persistent behavior at the "extinction → renewal" transition, supporting the interpretation that population activity during the choice period was related to behavior. The lack of a change in NCL population activity at the transition from persistent to exploratory behavior during extinction could be attributed to the fact that the behavioral protocol of pecking left or right still had a lot of similarity whereas the deliberate choice to discontinue the trial during avoidance was of a different quality.

In conclusion, our results demonstrate that NCL population activity changes during relevant task periods (i.e. stimulus presentation, choice period and outcome phase). These changes coincide with events of expectancy violation triggered via extinction or the renewal effect. On the one hand, our data indicated that some NCL neurons encoded an RPE signal already suggesting that the NCL population activity is directly related to the computation of the RPE. On the other hand, a subset of NCL neurons also seemed to represent reward as an absolute variable. Thus, the question remains what is the specific contribution of the NCL in the RPE network and how it relates back to decision-making. One potential account of prefrontal involvement in RPE coding was proposed by Wang et al. (2018). Their modelling results suggested that prefrontal areas such as the NCL provide temporally precise value-state updates of the CS-US association related to the learning process that are then propagated to midbrain dopamine neurons where the complete RPE signal is computed. In turn, the RPE signal is then projected back to prefrontal structures in a recurrent network to guide value-based action selection. Interestingly, this interpretation - based on mammalian datasets - is in full agreement with our results. The NCL population

displayed activity patterns consistent with a variety of input structures to the VTA, where subsets of neurons represented for example signals of pure reward, but also components of the RPE. These multifaceted components are then likely integrated in midbrain dopamine neurons to compute the complete RPE signal (Tian et al., 2016). Given that dopamine is a rather slow neuromodulator (for review see Seamans and Yang, 2004), it could be speculated that the NCL plays a role in providing temporal precise top-down controlled input to dopamine neurons reflecting the trial-wise update of the CS-US association as well as contextual information if behaviorally relevant. The notion that the NCL is involved in value-based action selection that could be guided via RPE signaling from dopamine neurons is also in accordance with the reported data as we found that the NCL was involved in decision-making at critical time points of the experiments. Thus, our results indicate that the RPE network seems to be comparable in birds and mammals. It is tempting to speculate that the underlying biological computations are conserved across species and appear to have limited degrees of freedom (Puig et al., 2014). Dopamine and prediction errors also influence learning in arthropods (e.g. Rohwedder et al., 2016; Terao and Mizunami, 2017), further supporting this assumption. This demonstrates the strength of RPEs as a learning signal as it seems to be a broadly distributed mechanism with similar underlying neural substrates across different animal classes to acquire and adapt behavior.

10. Methods

10.1. Subjects

Eight homing pigeons (*Columba livia*) that were acquired from local breeders were used as subjects. The birds were housed in individual wire-mesh cages located within a colony room that was controlled for temperature, humidity and the light/dark cycle (lights on from 08:00 am – 08:00 pm). The animals had ad libitum access to water and grit. Food access was restricted to experimental sessions on testing days and the animals were kept between 80 % - 90 % of their free-feeding body weight. The subjects were treated in accordance with the German guidelines for the care and use of animals in science and all procedures were approved by a national ethics committee of the State of North Rhine-Westphalia, Germany. All experimental conduct was in agreement with the European Communities Council Directive 86/609/EEC concerning the care and use of animals for experimental purposes.

10.2. Apparatus

Experiments were performed in a custom-built operant chamber (33 × 34 × 34 cm; (Packheiser et al., 2019b). Three horizontally aligned translucent response keys (5 × 5 cm) were located on the rear wall of the experimental chamber. For stimulus presentation, an LCD monitor was mounted against the rear wall. Successful pecks onto the response keys resulted in an audible feedback sound. A pellet feeder (http://www.jonasrose.net/open-labware/pellet-feeder/) was situated below the center key for reward delivery. In addition to food reward, correct responses were also indicated by an LED light beneath the feeding dish. Different sets of LED lights were affixed to the ceiling allowing for immediate changes of the contextual surrounding. The operant chamber was situated in a sound-attenuating cubicle to cancel out environmental noise. Furthermore, white noise (~60 dB) was played constantly during the experimental session to prevent distraction from external sources. The hardware was controlled by a custom MATLAB code using the Biopsy-Toolbox (The Mathworks, Natick, MA, USA; Rose et al., 2008).

10.3. Behavioral paradigm

Prior to the experimental procedure, the animals first received a shaping and then a pre-training. The shaping procedure was performed to habituate the animals to pecking onto all three response keys and was concluded once the animals reliably pecked onto each response key (>90 % responses). Then, the animals were initially pre-trained onto two stimuli that later served as familiar controls in the experiments. Furthermore, they also served as fix points for the animals during the experimental procedure as the animals did not have to learn their stimulus-response association in each individual session. One familiar stimulus was associated with the left response key and the other familiar stimulus was associated with the right response key. The stimulus-response association with the familiar stimuli was counterbalanced across animals. After the animals showed a consistently high performance rate for the familiar stimuli (>85 % correct responses in three consecutive sessions), the next pre-training step commenced. Here, two novel stimuli were introduced in each training session for which the animals had to learn to S-R association via trial and error. After the animals reliably acquired the S-R association of two novel stimuli within one session (>80 % correct responses computed as a running average over the last 100 trials in three consecutive sessions), the animals underwent surgery and were moved into the extinction paradigm.

In the experimental procedure, subjects were confronted with four different stimuli (two novel and two familiar stimuli) that were shown in a pseudorandomized order. Two of the stimuli required the animals to make a left choice whereas the other two stimuli required them to make a right choice in order to receive a food reward during the outcome period (Fig. 1A). The remaining two stimuli were unknown to the subjects prior to each experimental session for which the stimulus-response association had to be acquired during each recording session through trial and error. Thus, the acquisition phase of the experimental paradigm was identical to the last step of the pre-training procedure. The acquisition was conducted under house light conditions (context A) and followed the trial procedure as illustrated in Fig. 1B. In total, a minimum of 150 trials had to be completed to continue to the extinction phase. Furthermore, the animals had to initialize in 85 % of the trials and reach over 85 % correct responses for the novel stimuli and over 80 % correct responses for the familiar stimuli. These values were calculated as a running average over the past 100 trials to account for behavioral changes over time.

Each trial during the experimental procedure started with an initialization period during which an orange placeholder stimulus appeared in the center of the screen (Fig. 1B). The trial only continued if the animal successfully pecked on the center response key in time, otherwise the trial was aborted. Following initialization, the sample was presented for 2.5 s. The subjects were then required to acknowledge that they attended the sample by pecking on the orange placeholder stimulus once more in a confirmation period. After responding to the confirmation key, the placeholder stimulus disappeared and the two choice keys on the sides were illuminated. If the animals chose to peck on the correct choice key in the respective trial, a 2 s lasting reward period followed during which the animals received a food reward. Additionally, the pellet feeder was illuminated to highlight the reward delivery. If the animals chose to peck on the incorrect response key, the lights in the experimental chamber were shut off for 2 s to induce a mild punishment condition. Individual trials were separated by a 4 s long inter-trial interval (ITI).

The extinction phase differed in three key aspects from the acquisition phase (Fig. 1A and C). (1) One of the two novel stimuli was chosen at random to be extinguished meaning that it was neither followed by reward nor punishment after the animal made a choice. This stimulus is referred to as the extinction stimulus throughout the manuscript. The novel stimulus not chosen for extinction is referred to as the non-extinction stimulus. The outcome phase was replaced by a 2 s period void of any feedback. (2) Following the initialization peck, a colored LED light, the attribute of context B, replaced the white house light shown during the acquisition phase, which was the attribute of context A. Context B was represented by a red or green LED light in different experimental sessions to systematically vary the perceptual quality of the contextual information. The LED lights were present until the end of

the trial or the animal made an incorrect choice triggering the punishment condition. (3) To enhance contextual distinction between the acquisition context A and the extinction context B, we also changed the audio cues indicating correct or incorrect choices during extinction. To conclude the extinction phase, the pigeons had to initialize in a minimum of 85 % of the trials, deliver at least 80 % correct responses for the non-extinction stimulus and more than 75 % correct responses for the control stimuli. Performance for the extinction stimulus was required to drop below 20 % (formerly) correct responses. Thus, the animals were required to show avoidance behavior since exploratory behavior, i.e. alternating between the two choice keys, could have only resulted in a performance rate of around 50 %. All values were again calculated as a running average over the last 100 trials.

The renewal test was almost identical to the procedure during the acquisition as the contextual surrounding switched back to white house light conditions indicating the return to the acquisition context to the subjects (Fig. 1A and B). To measure the extent of the renewal effect, which is defined as the context dependent response recovery from extinction (Bouton, 2004), the extinction stimulus continued to be unfollowed by any feedback. The renewal test lasted for a fixed amount of 250 trials and did not require any behavioral criteria to be fulfilled in order to end the experiment.

10.4. Behavioral data analysis

Since the duration of acquisition and extinction were variable due to individual differences in reaching criteria, we divided these phases into six even blocks of trials for comparison. While the renewal test phase did not require a criterion to be reached, the animals sometimes stopped responding prior to performing all 250 trials. Therefore, we also divided this period into six blocks containing an even number of trials. We then calculated the number of conditioned responses and pecks onto the sample per block for each of the four presented stimuli. Pecking rates directly reflect the associated value of stimuli for pigeons (Kasties et al., 2016). Additionally, we calculated the number of avoidances during the extinction and the renewal test phase for all experimental stimuli. This avoidance behavior was constituted by an omission to respond to either the confirmation or choice key after the stimulus had been presented in the trial. Note that conditioned responses yielded a reward only during acquisition, but not during extinction and the renewal test phase.

For acquisition, a repeated measures ANOVA with the four experimental stimuli and training blocks as factors was performed to identify differences in conditioned responses and pecking rates across stimuli.

To measure the stability of the conditioned response for the novel stimuli after the acquisition to extinction transition, the last block of the acquisition and the first block of the extinction phase were compared using paired t-tests with conditioned choices and pecking rates as dependent variables. Furthermore, we quantified avoidance behavior during extinction. Here, a two-way ANOVA with the four experimental stimuli and training blocks as factors was performed to identify whether avoidance was exclusive for the extinction stimulus and whether it increased over time during extinction.

To understand the trial-by-trial dynamics of extinction learning on the behavioral level, we performed a more fine-tuned behavioral analysis for operant responses towards the extinction stimulus during the extinction phase. We computed a sliding window analysis to classify each individual trial into three distinct behavioral categories. The first behavioral category was labeled “persistent behavior” that was constituted by continuous conditioned responses on the formerly correct choice key. The second category was labeled “exploratory behavior” and was constituted by an alternating response between the two available choice keys. The last category was labeled “avoidance behavior” and was constituted by choice omissions after the extinction stimulus was presented. For categorization, behavioral responses were classified into conditioned responses (+1), choice omission (0) and alternative responses (-1) resulting in cumulative response functions (Gallistel et al., 2004). We then used a five-trial window size for the sliding window analysis to determine the category for each individual trial in which the cumulative responses were evaluated. To be classified as “persistent”, the sliding window starting from the first extinction trial had to contain at least three conditioned responses. For a trial to be classified as “avoidance”, the sliding window had to contain at least three avoidance trials. All trials that were neither classified as “persistent”, nor as “avoidance” were classified into the exploratory category. To identify behavioral changes over time, we assessed the trial during the extinction phase where the persistent, exploratory and avoidance behavior occurred for the first time during each session. Then, we computed the median of the first occurrences of these categories. We specifically did not use arithmetic means as they might have been skewed due to some sessions taking a long time for the animals to reach the criterion in the extinction phase. A median therefore likely represents a more typical value compared to an arithmetic mean.

We conducted paired t-tests between the performances in the last block of the extinction phase compared to each individual block of the renewal test to measure the extent of the renewal effect. The p-value was adjusted to $p < .008$ to account for multiple comparisons in accordance with a Bonferroni correction since the last block of the extinction was compared to all six blocks of the renewal test. Again, this analysis was identically conducted for pecking behavior.

10.5. Surgery

After the pre-training was completed, we implanted the subjects with custom built microdrives for electrophysiological recordings (Bilkey and Muir, 1999; Bilkey et al., 2003). The animals were anesthetized by a combinatory injection of Ketamine (Ketavet, 100 mg/mL; Zoetis, Germany) and Xylazine (Rompun, 20 mg/mL; Bayer, Germany) (0,065 mL per 100 g bodyweight in a 7:3 ratio). After injection, the feathers on the head were cut and the animals were subsequently placed in the stereotactic apparatus. A constant flow of Isoflurane (Forene, 100 % Isoflurane; Abbot, Germany) maintained the anesthesia throughout the surgery process. The scalp was incised and pulled sideways once the animals did no longer demonstrate any pain reflexes. Stainless steel screws were inserted into the skull to later serve as anchors for the dental cement. Above the coordinates for the NCL (AP + 6.0 mm, ML ±7.0 mm; Karten and Hodos, 1967), a small hole was drilled, and the dura mater was retracted. The electrodes were then inserted at the target location and fixated with dental cement. Another hole was drilled at the front of the skull for the placement of a silver wire that was melted at the tip (Teflon-coated silver wire, Ø =75 µm, Science Products, Hofheim, Germany) which served as ground electrode. The skin was then sutured and covered with antibiotics (Fucidine, 20 mg/g Natriumfusidat; Leo Pharma A/S, Denmark). Animals were treated with analgesics (Rimadyl, 50 mL/mL Carprofen; Zoetis, Germany) for three days and were allowed to recover for ten days following the surgery.

10.6. Electrophysiology

Physiological recordings were obtained by sixteen 40 µm formvar-insulated nichrome wires (impedances < 0.01 MΩ; California Fine Wire, Grover Beach, USA). Each electrode could serve as reference electrode during the recordings and was chosen online via visual inspection of the raw spike traces. 15 min prior to each recording session, electrodes were advanced by turning the screw attached to the microdrive at least one quarter revolution (~ 60 µm). Neural signals were amplified (1000 x) and band-pass filtered (0.3–3 kHz) using an extracellular recording amplifier (EXT-16DX amplifier, NPI electronics, Tamm, Germany). The data was then converted using an analog-to-digital converter at a sampling rate of 22 kHz and recorded using the software Spike2 (ADC-Converter: Power 1401−3; Spike2-Version: 8; Cambridge Electronic Design, Cambridge, UK). Raw spike traces were inspected offline for neural activity and digitally filtered (0.3–3 kHz) to

reduce movement related artifacts. Spikes were classified to be originating from a single unit using a custom-written MATLAB code (mlib toolbox, Maik Stüttgen, MATLAB central file exchange). Neural recordings had to fulfill the following criteria in order to be classified as single-units: (1) the signal-to-noise ratio was required to be at least two reflecting a deviation of signal to noise of at least eight standard deviations, (2) spike distributions of minimal and maximal amplitude peaks had to be normally distributed and (3) the interspike interval (< 4 ms) had to be void of spiking due to the refractory period following an action potential. Furthermore, units were controlled for motor related artifacts as the task involved pecking on response keys of the animals. Therefore, we manually inspected the channels for artifacts surrounding pecking events and computed peri-peck time histograms in an interval of ±20 ms around the pecks. Neurons that were especially active during this time period were excluded. The spike sorting process is described in detail in (Starosta et al., 2014).

10.7. Neural data analysis

For the neural analysis of the temporal dynamics of extinction learning, we used a sliding window analysis at the transition between experimental phases, i.e. between acquisition and extinction and between extinction and the renewal test. This type of analysis therefore compared firing rates between phases of learning throughout the experimental session (Fig. 3). For neural data analysis, we first normalized the raw firing rates via a z-transformation. For every trial of each experimental stimulus (two novel and two familiar), the mean firing rate calculated across the whole session in the ITI and the three experimental periods (μ) was subtracted from the raw firing rate within this time window in each individual trial (x_i) and divided by the according standard deviation (σ).

$$z = \frac{x_i - \mu}{\sigma}$$

The resulting z-scored firing rate provided a normalized and comparable measure of the neuronal response for each single unit. All further calculations were performed using normalized rather than raw firing rates.

For each individual neuron, we first analyzed firing rate changes in a time window within the ITI to establish a baseline condition of stochastic changes in neural activity. Here, we used a time window from -2000 ms to −1000 ms prior to the initialization onset of the current trial. This period was used to avoid both residual activity from the previous trial and anticipating activity from the next trial confounding the signal. To reliably assess neural activity patterns, we used a sliding window approach with a window size of 10 trials. We then determined a reference window (last 10 trials of the previous experimental phase) and a comparison window (e.g. the first 10 trials in the following experimental phase). Activity changes per cell were calculated according to the following formula:

$$Baseline\ \Delta_i^k = |\overline{C}_i^k - \overline{R}_i|$$

Here, $\overline{C}$ and $\overline{R}$ represent the average firing rates in a comparison window and the reference window during the ITI (see Fig. 3A). For each recorded neuron *i*, we calculated the difference *Baseline* Δ_i between $\overline{C}$ and $\overline{R}$. Since firing rate changes could be positive or negative after the phase transition, *Baseline* Δ_i was then transformed into absolute values. The transformation to absolute values was necessary due to the possible sign changes of RPE signals in the NCL as were found in the PFC (Assad & Eskandaar, 2011). In contrast to dopaminergic neurons, unexpected rewards do not necessarily elicit increased neural activity, nor do unexpected omissions of rewards necessarily elicit decreased neural activity. The analysis was conducted consecutively for every change in the starting trial of the comparison window *(k)* during extinction. The index k always increased by 1 so that the sliding window moved forward in a trial-by-trial fashion. We then repeated the identical analysis for all cells in the three experimental periods, namely the sample presentation, the choice period and the outcome period using the same formula was for the calculation of the baseline condition:

$$Exp\ \Delta_i^k = |\overline{C}_i^k - \overline{R}_i|$$

Sample presentation activity was measured from stimulus onset until the end of the stimulus presentation. As the animals started to avoid making a choice during the extinction phase, the choice and the outcome period were estimated based on the mean choice time within each session. Please note that the estimation was performed for all experimental stimuli and not just for the extinction stimulus to keep the analysis consistent. Therefore, coding differences between experimental stimuli could not be attributed to the analysis. Choice-related activity was thus computed from 2000 ms after stimulus onset to the mean choice time of the session. The interval to investigate the reward-related activity was computed from the mean choice time until 2000 ms later (duration of the outcome period).

In a final step, we calculated the net activity change in the population code by comparing the degree of change during the baseline period of all neurons to the degree of change in each experimental period for all neurons across all comparison windows. Here, paired *t*-tests and measures of effect size (Cohen's d) were used to quantify the strength of the activity change for the whole population. Therefore, we could determine a discrete value of activity change for the whole population for each comparison window (Fig. 3B). To determine a meaningful threshold for significant activity changes on the population level, a permutation test was conducted in which the experimental data was shuffled 1000 times to identify how often these changes occur by chance (SI Fig. 4). A value of Cohen's d ≥ 0.3 was determined to be a conservative threshold.

For the phase transition from acquisition to extinction, population changes in activity were quantified for 30 consecutive comparison windows for each experimental stimulus and each experimental period. Additionally, we compared activity changes during extinction for the choice period by comparing the last 10 trials before the onset of exploratory behavior to the first 10 trials after the onset of exploratory behavior taking into account the session to session variability in behavioral switches. The identical analysis was conducted at the switch from exploratory to avoidance behavior (reference window = last 10 trials before avoidance onset, comparison window = first 10 trials after avoidance onset). The test phase was only analyzed until the 10th comparison window. The reason for this was twofold: first, renewal is a rather short-lived phenomenon and therefore could only properly analyzed in the early stages after the contextual switch (see SI Fig. 1 for example). Second, there was no behavioral criterion for the animals to reach to complete the test phase. Thus, many sessions ended prematurely due to the animals' cessation in responding and there were no more trials to analyze for a subset of neurons.

For outcome period comparisons specifically at the phase transition from acquisition to extinction, we also investigated sign changes of individual neurons exhibiting signal changes of an effect size of Cohen's d > 1. If the average firing rate in the last 10 trials of acquisition was higher compared to the average firing rate of the neuron overall and the average firing rate of the first 10 trials of extinction was lower compared to the average firing rate of the neuron, a neuron was classified as a sign changing neuron. Sign changing neurons were additionally investigated for decreasing activity by comparing activity levels in the first and last comparison window during extinction. Finally, we investigated whether some neurons exhibited continuous activity changes during the outcome period. A neuron was classified as such if it demonstrated an activity change of at least Cohen's d > 0.5 during each individual comparison window.

11. Supplementary information

11.1. Supplementary results

Overall, the left novel stimulus was extinguished in 22 sessions, and the right novel stimulus in 34 sessions.

11.2. Acquisition (Context A)

To identify learning differences between familiar and novel stimuli, we first compared performance rates for the four experimental stimuli during acquisition. Here, we found that novel stimuli received less conditioned choices compared to the familiar ones in the first five training blocks ($F_{(3,825)} = 10.49$, $p < .001$, $\eta_p^2 = 0.16$, repeated measures ANOVA; all p's $< .007$, Fig. 2A, left panel). By the sixth acquisition block no observable difference in choice behavior between novel and familiar stimuli could be detected, indicating that S-R associations for the novel stimuli had been fully learned. Both novel stimuli were learned at equal pace as there was no difference in conditioned choices between them during training blocks ($p > .118$).

We then analyzed avoidance behavior for all experimental stimuli which occurs regularly during extinction learning (Packheiser et al., 2019a). Responses were classified as avoidance behavior whenever a choice was omitted after the stimulus had been presented to the animal, i.e. either during the confirmation or choice period. Presentations of the extinction stimulus were followed by a significantly higher rate of avoidances compared to the other experimental stimuli ($F_{(3,165)} = 254.38$, $p < .001$, $\eta_p^2 = 0.82$; extinction stimulus mean = 55.9 % avoidance; non-extinction stimulus mean = 2.5 % avoidance, control left mean = 1.7 % avoidance, control right mean = 1.0 % avoidance, all p's $< .001$). Thus, the pigeons selectively discontinued trials in which the extinction stimulus was presented. Furthermore, the increase of avoidance correlated significantly with the decrease of pecking onto the stimulus ($r = -0.68$, $p < .001$). Since pecking rate is, in turn, an indicator of associative strength (Kasties et al., 2016), our results imply that avoidance might be inversely correlated with the associative strength of the extinction stimulus. We also found that pecks onto the right control stimulus were increased compared to the left control stimulus during extinction ($p < .001$). A possible explanation could relate to asymmetry of visual object categorization ability in birds with a superiority of the left hemisphere or right eye (Yamazaki et al., 2007). The uncertainty induced during extinction might have brought forth such a bias in our experiment as well.

Since the extinction stimulus continued to be unrewarded in the test phase, the conditioned response was subsequently re-extinguished. To measure re-extinction in the test phase, we calculated the increase of avoidance behavior over time for the extinction stimulus. Here, the proportion of avoidance significantly increased from the first to the fourth block in the test phase (($F_{(3,165)} = 35.30$, $p < .001$, $\eta_p^2 = 0.39$), all p's $< .023$). Again, avoidance showed a negative correlation with the number of pecks onto the extinction stimulus ($r = -0.62$, $p < .001$).

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projectnumber 316803389 SFB 1280 (projects A01 and F01), and the Research Training Group "Situated Cognition" (GRK 2185/1).

Appendix A. The Peer Review Overview and Supplementary data

The Peer Review Overview and Supplementary data associated with this article can be found in the online version: https://doi.org/10.1016/j.pneurobio.2020.101901.

References

Asaad, W.F., Eskandar, E.N., 2011. Encoding of both positive and negative reward prediction errors by neurons of the primate lateral prefrontal cortex and caudate nucleus. J. Neurosci. 31 (49), 17772–17787. https://doi.org/10.1523/JNEUROSCI.3793-11.2011.

Bayer, H.M., Glimcher, P.W., 2005. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47 (1), 129–141. https://doi.org/10.1016/j.neuron.2005.05.020.

Bilkey, D.K., Muir, G.M., 1999. A low cost, high precision subminiature microdrive for extracellular unit recording in behaving animals. J. Neurosci. Methods 92 (1-2), 87–90. https://doi.org/10.1016/S0165-0270(99)00102-8.

Bilkey, D.K., Russell, N., Colombo, M., 2003. A lightweight microdrive for single-unit recording in freely moving rats and pigeons. Methods 30 (2), 152–158. https://doi.org/10.1016/S1046-2023(03)00076-8.

Bouton, M.E., 2004. Context and behavioral processes in extinction. Learn. Mem. 11 (5), 485–494. https://doi.org/10.1101/lm.78804.

Coddington, L.T., Dudman, J.T., 2018. The timing of action determines reward prediction signals in identified midbrain dopamine neurons. Nat. Neurosci. 21 (11), 1563–1573. https://doi.org/10.1038/s41593-018-0245-7.

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.

de Voogd, L.D., Murray, Y.P.J., Barte, R.M., van der Heide, A., Fernández, G., Doeller, C. F., Hermans, E.J., 2020. The role of hippocampal spatial representations in contextualization and generalization of fear. NeuroImage 206, 116308. https://doi.org/10.1016/j.neuroimage.2019.116308.

Eckerman, D.A., Lanson, R.N., 1969. Variability of response location for pigeons responding under continuous reinforcement, intermittent reinforcement, and extinction. J. Exp. Anal. Behav. 12 (1), 73–80. https://doi.org/10.1901/jeab.1969.12-73.

Enomoto, K., Matsumoto, N., Nakai, S., Satoh, T., Sato, T.K., Ueda, Y., et al., 2011. Dopamine neurons learn to encode the long-term value of multiple future rewards. Proc. Natl. Acad. Sci. USA 108 (37), 15462–15467. https://doi.org/10.1073/pnas.1014457108.

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

Gallistel, C.R., Fairhurst, S., Balsam, P., 2004. The learning curve: implications of a quantitative analysis. Proc. Natl. Acad. Sci. USA 101 (36), 13124–13131. https://doi.org/10.1073/pnas.0404965101.

Gao, M., Lengersdorf, D., Stüttgen, M.C., Güntürkün, O., 2019. Transient inactivation of the visual-associative nidopallium frontolaterale (NFL) impairs extinction learning and context encoding in pigeons. Neurobiol. Learn. Mem. 158, 50–59. https://doi.org/10.1016/j.nlm.2019.01.012.

Gluck, M.A., Bower, G.H., 1988. From conditioning to category learning: an adaptive network model. J. Exp. Psychol. Gen. 117 (3), 227–247. https://doi.org/10.1037/0096-3445.117.3.227.

Güntürkün, O., 2005. The avian 'prefrontal cortex' and cognition. Curr. Opin. Neurobiol. 15 (6), 686–693. https://doi.org/10.1016/j.conb.2005.10.003.

Hollerman, J.R., Schultz, W., 1998. Dopamine neurons report an error in the temporal prediction of reward during learning. Nat. Neurosci. 1 (4), 304–309. https://doi.org/10.1038/1124.

Karten, H.J., Hodos, W., 1967. Stereotaxic Atlas of the Brain of the Pigeon (Columba Livia). Johns Hopkins Press, Baltimore.

Kasties, N., Starosta, S., Güntürkün, O., Stüttgen, M.C., 2016. Neurons in the pigeon caudolateral nidopallium differentiate Pavlovian conditioned stimuli but not their associated reward value in a sign-tracking paradigm. Sci. Rep. 6, 35469. https://doi.org/10.1038/srep35469.

Kinloch, J.M., Foster, T.M., McEwan, J.S.A., 2009. Extinction-induced variability in human behavior. Psychol. Rec. 59 (3), 347–369. https://doi.org/10.1007/BF03395669.

Kobayashi, S., Schultz, W., 2008. Influence of reward delays on responses of dopamine neurons. J. Neurosci. 28 (31), 7837–7846. https://doi.org/10.1523/JNEUROSCI.1600-08.2008.

Kröner, S., Güntürkün, O., 1999. Afferent and efferent connections of the caudolateral neostriatum in the pigeon (*Columba livia*): a retro- and anterograde pathway tracing study. J. Comp. Neurol. 407 (2), 228–260. https://doi.org/10.1002/(SICI)1096-9861(19990503)407:2<228::AID-CNE6>3.0.CO;2-2.

Lak, A., Stauffer, W.R., Schultz, W., 2016. Dopamine neurons learn relative chosen value from probabilistic rewards. ELife 5. https://doi.org/10.7554/eLife.18044.

Lengersdorf, D., Pusch, R., Güntürkün, O., Stüttgen, M.C., 2014a. Neurons in the pigeon nidopallium caudolaterale signal the selection and execution of perceptual decisions. Eur. J. Neurosci. 40 (9), 3316–3327. https://doi.org/10.1111/ejn.12698.

Lengersdorf, D., Stüttgen, M.C., Uengoer, M., Güntürkün, O., 2014b. Transient inactivation of the pigeon hippocampus or the nidopallium caudolaterale during

extinction learning impairs extinction retrieval in an appetitive conditioning paradigm. Behav. Brain Res. 265, 93–100. https://doi.org/10.1016/j.bbr.2014.02.025.
Lengersdorf, D., Marks, D., Uengoer, M., Stüttgen, M.C., Güntürkün, O., 2015. Blocking NMDA-receptors in the pigeon's "prefrontal" caudal nidopallium impairs appetitive extinction learning in a sign-tracking paradigm. Front. Behav. Neurosci. 9, 85. https://doi.org/10.3389/fnbeh.2015.00085.
Lissek, S., Glaubitz, B., Schmidt-Wilcke, T., Tegenthoff, M., 2016. Hippocampal context processing during acquisition of a predictive learning task is associated with renewal in extinction recall. J. Cogn. Neurosci. 28 (5), 747–762. https://doi.org/10.1162/jocn_a_00928.
Matsumoto, M., Hikosaka, O., 2009. Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature 459 (7248), 837–841. https://doi.org/10.1038/nature08028.
Menegas, W., Babayan, B.M., Uchida, N., Watabe-Uchida, M., 2017. Opposite initialization to novel cues in dopamine signaling in ventral and posterior striatum in mice. ELife 6. https://doi.org/10.7554/eLife.21886.
Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167.
Mirenowicz, J., Schultz, W., 1994. Importance of unpredictability for reward responses in primate dopamine neurons. J. Neurophysiol. 72 (2), 1024–1027. https://doi.org/10.1152/jn.1994.72.2.1024.
Morris, G., Nevet, A., Arkadir, D., Vaadia, E., Bergman, H., 2006. Midbrain dopamine neurons encode decisions for future action. Nat. Neurosci. 9 (8), 1057–1063. https://doi.org/10.1038/nn1743.
Neuringer, A., Kornell, N., Olufs, M., 2001. Stability and variability in extinction. J. Exp. Psychol. Anim. Behav. Process. 27 (1), 79–94. https://doi.org/10.1037/0097-7403.27.1.79.
Orsini, C.A., Maren, S., 2012. Neural and cellular mechanisms of fear and extinction memory formation. Neurosci. Biobehav. Rev. 36 (7), 1773–1802. https://doi.org/10.1016/j.neubiorev.2011.12.014.
Oya, H., Adolphs, R., Kawasaki, H., Bechara, A., Damasio, A., Howard, M.A., 2005. Electrophysiological correlates of reward prediction error recorded in the human prefrontal cortex. Proc. Natl. Acad. Sci. USA 102 (23), 8351–8356. https://doi.org/10.1073/pnas.0500899102.
Packheiser, J., Güntürkün, O., Pusch, R., 2019a. Renewal of extinguished behavior in pigeons (*Columba livia*) does not require memory consolidation of acquisition or extinction in a free-operant appetitive conditioning paradigm. Behav. Brain Res. 370, 111947 https://doi.org/10.1016/j.bbr.2019.111947.
Packheiser, J., Pusch, R., Stein, C.C., Güntürkün, O., Lachnit, H., Uengoer, M., 2019b. How competitive is cue competition? Q. J. Exp. Psychol. (2006), 1747021819866967 https://doi.org/10.1177/1747021819866967.
Pan, W.X., Brown, J., Dudman, J.T., 2013. Neural signals of extinction in the inhibitory microcircuit of the ventral midbrain. Nat. Neurosci. 16 (1), 71–78. https://doi.org/10.1038/nn.3283.
Puig, M.V., Rose, J., Schmidt, R., Freund, N., 2014. Dopamine modulation of learning and memory in the prefrontal cortex: insights from studies in primates, rodents, and birds. Front. Neural Circuits 8, 93. https://doi.org/10.3389/fncir.2014.00093.
Rangel, A., Camerer, C., Montague, P.R., 2008. A framework for studying the neurobiology of value-based decision making. Nat. Rev. Neurosci. 9 (7), 545–556. https://doi.org/10.1038/nrn2357.
Rescorla, R.A., Wagner, A.R., 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. Classic. Conditioning II: Curr. Res. Theory 2, 64–99.
Rick, J.H., Horvitz, J.C., Balsam, P.D., 2006. Dopamine receptor blockade and extinction differentially affect behavioral variability. Behav. Neurosci. 120 (2), 488–492. https://doi.org/10.1037/0735-7044.120.2.488.
Rohwedder, A., Wenz, N.L., Stehle, B., Huser, A., Yamagata, N., Zlatic, M., et al., 2016. Four Individually Identified Paired Dopamine Neurons Signal Reward in Larval Drosophila. Curr. Biol. 26 (5), 661–669. https://doi.org/10.1016/j.cub.2016.01.012.
Rose, J., Otto, T., Dittrich, L., 2008. The Biopsychology-Toolbox: a free, open-source Matlab-toolbox for the control of behavioral experiments. J. Neurosci. Methods 175 (1), 104–107. https://doi.org/10.1016/j.jneumeth.2008.08.006.
Sadacca, B.F., Jones, J.L., Schoenbaum, G., 2016. Midbrain dopamine neurons compute inferred and cached value prediction errors in a common framework. ELife 5. https://doi.org/10.7554/eLife.13665.
Saddoris, M.P., Sugam, J.A., Stuber, G.D., Witten, I.B., Deisseroth, K., Carelli, R.M., 2015. Mesolimbic dopamine dynamically tracks, and is causally linked to, discrete aspects of value-based decision making. Biol. Psychiatry 77 (10), 903–911. https://doi.org/10.1016/j.biopsych.2014.10.024.
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. https://doi.org/10.7554/eLife.38818.
Schultz, W., 2007. Behavioral dopamine signals. Trends Neurosci. 30 (5), 203–210. https://doi.org/10.1016/j.tins.2007.03.007.
Schultz, W., 2015. Neuronal reward and decision signals: from theories to data. Physiol. Rev. 95 (3), 853–951. https://doi.org/10.1152/physrev.00023.2014.
Schultz, W., 2016a. Dopamine reward prediction-error signalling: a two-component response. Nat. Rev. Neurosci. 17 (3), 183–195. https://doi.org/10.1038/nrn.2015.26.
Schultz, W., 2016b. Dopamine reward prediction error coding. Dialogues Clin. Neurosci. 18 (1), 23–32.
Schultz, W., Dickinson, A., 2000. Neuronal coding of prediction errors. Annu. Rev. Neurosci. 23, 473–500. https://doi.org/10.1146/annurev.neuro.23.1.473.
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.
Seamans, J.K., Yang, C.R., 2004. The principal features and mechanisms of dopamine modulation in the prefrontal cortex. Prog. Neurobiol. 74 (1), 1–58. https://doi.org/10.1016/j.pneurobio.2004.05.006.
Slopsema, J.S., van der Gugten, J., de Bruin, J.P.C., 1982. Regional concentrations of noradrenaline and dopamine in the frontal cortex of the rat: dopaminergic innervation of the prefrontal subareas and lateralization of prefrontal dopamine. Brain Res. 250 (1), 197–200. https://doi.org/10.1016/0006-8993(82)90970-2.
Starosta, S., Stüttgen, M.C., Güntürkün, O., 2014. Recording single neurons' action potentials from freely moving pigeons across three stages of learning. J. Vis. Exp. (88), e51283. https://doi.org/10.3791/51283.
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. Nat. Neurosci. 16 (7), 966–973. https://doi.org/10.1038/nn.3413.
Sutton, R.S., Barto, A.G., 1998. Introduction to Reinforcement Learning, Vol. 2. MIT press., Cambridge.
Terao, K., Mizunami, M., 2017. Roles of dopamine neurons in mediating the prediction error in aversive learning in insects. Sci. Rep. 7 (1), 14694. https://doi.org/10.1038/s41598-017-14473-y.
Tian, J., Huang, R., Cohen, J.Y., Osakada, F., Kobak, D., Machens, C.K., et al., 2016. Distributed and mixed information in monosynaptic inputs to dopamine neurons. Neuron 91 (6), 1374–1389. https://doi.org/10.1016/j.neuron.2016.08.018.
Veit, L., Nieder, A., 2013. Abstract rule neurons in the endbrain support intelligent behaviour in corvid songbirds. Nat. Commun. 4, 2878. https://doi.org/10.1038/ncomms3878.
Veit, L., Hartmann, K., Nieder, A., 2014. Neuronal correlates of visual working memory in the corvid endbrain. J. Neurosci. 34 (23), 7778–7786. https://doi.org/10.1523/JNEUROSCI.0612-14.2014.
Veit, L., Pidpruzhnykova, G., Nieder, A., 2015. Associative learning rapidly establishes neuronal representations of upcoming behavioral choices in crows. Proc. Natl. Acad. Sci. USA 112 (49), 15208–15213. https://doi.org/10.1073/pnas.1509760112.
Walker, S.C., Robbins, T.W., Roberts, A.C., 2009. Differential contributions of dopamine and serotonin to orbitofrontal cortex function in the marmoset. Cereb. Cortex 19 (4), 889–898. https://doi.org/10.1093/cercor/bhn136.
Wang, J., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J.Z., et al., 2018. Prefrontal cortex as a meta-reinforcement learning system. Nat. Neurosci. 21 (6), 860–868. https://doi.org/10.1038/s41593-018-0147-8.
Watabe-Uchida, M., Eshel, N., Uchida, N., 2017. Neural circuitry of reward prediction error. Annu. Rev. Neurosci. 40, 373–394. https://doi.org/10.1146/annurev-neuro-072116-031109.
Wynne, B., Güntürkün, O., 1995. Dopaminergic innervation of the telencephalon of the pigeon (*Columba livia*): a study with antibodies against tyrosine hydroxylase and dopamine. J. Comp. Neurol. 357 (3), 446–464. https://doi.org/10.1002/cne.903570309.
Yamazaki, Y., Aust, U., Huber, L., Hausmann, M., Güntürkün, O., 2007. Lateralized cognition: asymmetrical and complementary strategies of pigeons during discrimination of the "human concept". Cognition 104 (2), 315–344. https://doi.org/10.1016/j.cognition.2006.07.004.