A retrospective and stepwise learning strategy revealed by neuronal activity in the basal forebrain

Hachi E. Manzur[1], Ksenia Vlasov[1], Shih-Chieh Lin[1,2,3]*

[1.] Neural Circuits and Cognition Unit, Laboratory of Behavioral Neuroscience, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA

[2.] Institute of Neuroscience, National Yang Ming Chiao Tung University, Taipei, Taiwan

[3.] Brain Research Center, National Yang Ming Chiao Tung University, Taipei, Taiwan

*Correspondence to: shihchieh.lin@nycu.edu.tw

Abstract

Associative learning is a fundamental cognitive capacity that allows animals and humans to learn the predictive relationship between behavioral events and rewarding outcomes. While the process of learning is commonly conceptualized as a prospective strategy (learning how behavioral events predict *future* rewards), here we provide behavioral and neurophysiological evidence to show that animals may instead employ a retrospective and stepwise learning strategy (learning how the reward is predicted by *preceding* behavioral events). In rats learning a new association in which the reward was paired with a sequence of behavioral events, learning started from the event closest to the reward and sequentially incorporated earlier events into animals' internal model. The learning of each behavioral event as a new reward predictor was accompanied by the emergence of basal forebrain (BF) neuronal responses toward that event. BF activities quantitatively conveyed a reward prediction error signal associated with the behavioral event, and promoted reward-seeking behavioral sequences containing the newly learned event. As the internal model incorporated more behavioral events as reward predictors, non-rewarded behavioral sequences that were once compatible with the internal model during early stages of learning became incompatible and were sequentially eliminated. Together, these results demonstrate how the retrospective and stepwise learning strategy can effectively establish animals' internal model during the learning process and lead to the sequential refinement of reward-seeking behaviors. These results also highlight the functional significance of BF neuronal activities, which provided unique insights into the covert dynamics of the learning process in single trials.

Introduction

Associative learning is essential for survival and allows animals and humans to predict future reward based on behavioral events, such as environmental stimuli[1,2] or their own actions[3,4]. Understanding the algorithmic principles of associative learning has been a central question in psychology and neuroscience[5–12], and has broad implications in machine learning[13] and artificial intelligence[14]. Much evidence supports the view that animals form internal models of associative relationships between behavioral events and rewarding outcomes[5–8]. However, it remains unclear how internal reward-prediction models are established in the first place, and how such models dynamically evolve during the early phase of new associative learning.

The learning strategies for building the internal model can be conceptualized under two broad categories: prospective[8,9] and retrospective strategies[15,16]. Under the prospective strategy, animals learn about how the current behavioral state predicts events and rewards in the future. This forward-looking prospective strategy is widely used in well-learned contexts as an effective means to guide decision making through the prediction of future rewards. An alternative, but not mutually exclusive, strategy that has been proposed recently is the backward-looking retrospective strategy[15,16]. Under the retrospective strategy, learning starts from the rewarded state and animals learn about how the rewarded state is predicted by preceding behavioral events. Recent studies support that animals do learn about retrospective associative relationships between behavioral events[17], and the retrospective strategy might be especially powerful during the learning process for discovering the connection between rare rewards with preceding behavioral events[15]. The current study seeks to provide empirical support for the use of retrospective strategy during the learning process.

To investigate whether animals use the retrospective strategy to build their internal reward-prediction models during the learning process, an ideal setting to test this idea is when animals must learn the association between the reward and a sequence of behavioral events. Under this setting, the retrospective strategy predicts that learning would be triggered by the unexpected reward and animals would first learn about the last event closest to the reward as the reward predictor, and subsequently learn about earlier behavioral events. Moreover, as the internal model expands and sequentially incorporates more behavioral events, behavioral responses should evolve accordingly and follow specific sequences. The current study tested these predictions and sought to identify the behavioral signatures of the retrospective learning strategy.

In order to provide an independent validation for the behavioral observations, we further focused on a learning-related neural activity in the basal forebrain[18–24] (BF), which conveys a reward-prediction error signal associated with rewards and reward-predicting behavioral events[23–27]. Specifically, we focused on a special subset of noncholinergic BF neurons, which will be referred to as BF bursting neurons[18–24], that show highly robust phasic bursting responses to reward-predicting sensory stimuli irrespective of their sensory modalities. Moreover, such responses only emerge after reward-based associative learning[23], and are tightly coupled with behavioral performance and promotes faster decision speeds[22–24]. By observing the temporal evolution of BF activities throughout the learning process, we will be able to determine which behavioral events animals have learned to predict reward, and when such learning takes place. In the context of retrospective learning strategy, we predict that BF

responses should first emerge toward the last behavioral event closest to the reward, and subsequently develop toward the earlier behavioral events. In other words, the temporal evolution of BF activities should mirror the behavioral learning process, and reveal the dynamic expansion of the internal reward-prediction model during the learning process.

Results

To better understand the strategy the animals adopt during associative learning, we first trained adult Long-Evans rats in an auditory discrimination task. Rats entered the fixation port to initiate each trial, where they encountered three trial types (S^{left}; S^{right}; catch) with equal probabilities that respectively indicated sucrose water reward in the left or right port, or no reward in the case of catch trials (no stimulus) (Figure 1A). During the initial auditory discrimination phase, S^{left} and S^{right} were two distinct sound stimuli. After reaching asymptotic performance, rats entered the new learning phase (denoted as the D_0 session), in which the S^{right} stimulus was switched to a novel light stimulus that rats never encountered before (Figure 1B, Table 1). This enabled rats to learn the new stimulus-reward association while maintaining stable levels of performance toward the previously-learned S^{left} sound stimulus (Figure 1C). All rats (N=7) began responding in the new light trials within the first three sessions, and reached >90% correct responses within fourth sessions of new learning and maintained stable performance afterwards (Figure 1C, Table 2).

Detailed examination of behavioral response patterns, however, revealed that the new learning experience not only changed behaviors in the new light trials, but also led to the emergence of two other types of behaviors toward the right reward port: catch licks and no-fixation licks (Figure 1C). Catch licks refers to licking responses toward the right reward port in catch trials when no sensory stimuli were presented. No-fixation licks refers to the situation in which rats directly licked at the right reward port without properly initiating the trial by nosepoking in the center fixation port. These two types of behaviors were largely absent before the new learning phase.

(Figure 1. New associative learning led to sequential refinements of reward-seeking behaviors, consistent with a retrospective learning strategy)

The emergence of no-fixation licks and catch licks in the new learning phase was curious because both behaviors were not rewarded. Moreover, the frequency of both behaviors peaked during early sessions of the new learning phase, and subsequently diminished in later sessions (Figure 1C). Their temporal sequence was highly consistent across animals despite individual behavioral variabilities: the peak of no-fixation licks occurred earlier than the peak of catch licks in most animals (N=6/7) or in the same session (N=1/7). We will denote the peak of no-fixation licks as the D_1 session, and the peak of catch licks as the D_2 session in each animal. The D_1 and D_2 sessions allowed us to identify similar learning stages across animals despite their individual differences in learning dynamics.

These behavioral observations indicate that the new learning not only induced behavioral changes in the new light trials that were rewarded, but also led to the emergence of non-rewarded licking behaviors toward the same reward port. Both the rewarded and non-rewarded licking behaviors all emerged in the D_1 session (Figure 1C, middle panel, no response in the D_1-1 session), suggesting that the D_1 session was likely when the initial learning took place.

Based on these observations, new learning could be characterized as the sequential refinement of reward-seeking behaviors toward the right reward port, in which non-rewarded licking behaviors were sequentially eliminated until licking was selectively present in the new light trial.

The overall pattern of reward-seeking behaviors toward the right reward port was not easily compatible with the prospective learning strategy of decision making (Figure 1D). In the prospective model, the three types of rightward licking behaviors were scattered at different branches of the decision tree, which made it difficult to explain the conserved pattern of sequential refinement. There was also little evidence that animals explored other branches of the decision trees during learning.

On the other hand, the retrospective learning strategy provided a simple and parsimonious model that directly linked the three types of rightward licking behaviors in sequential orders (Figure 1E). We hypothesize that the sequential refinement of reward-seeking behaviors might result from the stepwise expansion of the internal model under the retrospective learning strategy. During the initial stage of learning, the internal model would only incorporate the behavioral event closest to the reward (lick right) as the reward predictor. Since all three types of rightward licking behaviors contained this behavioral event, and therefore would all predict reward, animals would engage in all three types of reward-seeking behaviors at this initial stage of learning. As the internal model subsequently expanded and incorporated earlier events (fix-out and then the light stimulus), no-fixation licks and catch licks would become incompatible with the expanded internal model because they did not contain all the reward-predicting events. As a result, no-fixation licks and catch licks would no longer predict reward in later stages of learning, and would be subsequently eliminated. We suggest that the sequential refinement of reward-seeking behaviors might represent the behavioral signature of the retrospective and stepwise learning strategy.

(Figure 2. Abrupt transition in reward-seeking behaviors corresponded to increased neuronal activity in the BF during initial learning)

To validate the behavioral observations and understand the underlying neural dynamics, we recorded BF neuronal activity throughout the learning process (Figure 2A) and used the consistent S^{left} sound as the control stimulus to identify stable populations of BF bursting neurons (Figure 2B, Figure S1). A total of 1453 BF single units were recorded over 45 sessions (N=7 rats), of which 70% (1013/1453) were classified as BF bursting neurons based on their stereotypical phasic response to the S^{left} sound (22.5±7.3 neurons per session, mean±std) (Figure 2B). The population response of BF bursting neurons were highly consistent across animals (Figure 2C), and remained remarkably stable in S^{left} trials throughout the learning process (Figure 2D). The inclusion of S^{left} trials therefore allowed us to record from stable and representative populations of BF bursting neurons throughout the learning process, and to investigate how their activities dynamically evolved in other trial types during learning at single trial resolution (Figure 2E).

We first applied this approach to understand the behavioral and neural dynamics in the D_1 session because all three types of rightward licks emerged in this session (Figure 1C, middle panel). Detailed analysis of behavioral responses from a representative session (Figure 2E) revealed that the three types of rightward reward-seeking behaviors in fact emerged abruptly after a transition point (see Methods for definition). Rightward licking behaviors were mostly

absent before the transition point, and rapidly switched to almost 100% licking after the transition point. This pattern was consistently observed across all animals (Figure 3A1).

At the neuronal level, there was a corresponding increase in the activity of BF bursting neurons that rapidly emerged after the abrupt transition in reward-seeking behaviors (Figure 2E, 3B1). This increase in BF activity was most prominent in the epoch before the trial outcome (Figure 2E2). In contrast, in trials before the transition point, BF bursting neurons did not show similar activity increases in the corresponding time window after exiting the fixation port (Figure 2E, 3B1). We will refer to the BF activity in this window as the BF evaluation response (see Methods for definition) because it reflected animals' internal evaluation when no additional sensory stimuli were presented during this epoch .

(Figure 3. Unexpected reward triggered the rapid emergence and subsequent refinement of reward-seeking behaviors and BF activity)

We further examined the respective contributions of the three types of rightward licking behaviors during the transition process in D_1 sessions (Figure 3). We noted that the first trial after the behavioral transition was typically a rewarded light lick trial (Figure 3A2 inset). This observation is consistent with the retrospective learning strategy, which predicts that learning would be triggered by the receipt of an unexpected reward.

Immediately after the abrupt transition, all three types of rightward licking behaviors emerged (Figure 3A). No-fixation licks, in particular, were most frequently observed within the first 60 trials after the transition point, before declining subsequently in the next 60 trials and occurred less frequently than light licks and catch licks (Figure 3A3).

At the neuronal activity level, BF evaluation responses quickly increased after the transition point in all three types of rightward licking behaviors (Figure 3B). The amplitudes of BF evaluation responses were similar between the three types of rightward licking behaviors within the first 60 trials after the transition. Subsequently, the BF evaluation response in no-fixation licks declined in the next 60 trials, relative to the other two types of rightward licking behaviors (Figure 3B2-3).

These results support that the first step of retrospective learning occurred at the transition point in the D_1 session (Figure 1E): the unexpected reward triggered the learning of the preceding state, i.e. licking toward the right reward port, as the reward predictor and led to the abrupt behavioral transition. During this initial stage (i.e. the first 60 trials after the transition point in the D_1 session), all three types of rightward licking behaviors contained the right lick event and were therefore compatible with this internal model. BF activity was elevated whenever animals approached and licked the right reward port regardless of whether they had exited from the center fixation port.

At roughly 60 trials after the transition point, the second step of the retrospective learning took place and animals expanded their internal model and incorporated a second reward predictor: exiting the fixation port (Figure 1E). As a result, no-fixation licks were no longer compatible with this model (because they did not contain the fixation exit event), resulting in diminished BF evaluation responses and the elimination of this behavior. The other two types of rightward

licks, light licks and catch licks, remained compatible with this expanded model and maintained high levels of BF evaluation responses.

(Figure 4. BF neurons did not respond to the new light stimulus during initial learning)

The similarities between light licks and catch licks in the D_1 session, at both behavioral and neural activity levels (Figure 3), suggested that animals likely treated these two trial types as the same during the D_1 session. Such a possibility would predict that the new light stimulus was not the cause of reward-seeking behavior in the D_1 session and would not elicit responses in BF bursting neurons.

We tested this prediction by comparing BF activities between light and catch trials, as well as between lick and no lick trials, in D_1 sessions (Figure 4). Indeed, BF activities in light trials were highly similar to those in catch trials, in the epochs before exiting the fixation port. This was true regardless of whether animals subsequently licked at the right reward port. This observation confirmed the prediction that the new light stimulus did not activate BF bursting neurons in the D_1 session, despite the near-perfect behavioral performance in light trials after the transition point.

In contrast, after fixation port exit, there were similar increases in BF activity when animals licked at the right reward port in both light and catch trials (Figure 4), which corresponded to the BF evaluation response described earlier (Figure 3B). Increases in BF activity after fixation port exit reliably distinguished between lick and no lick trials, but not between light and catch trials (Figure 4C). These observations support the idea that light and catch trials were treated as the same in the D_1 session, and that the light stimulus was not used as a reward predictor at this stage of learning.

(Figure 5. BF responses to light onset emerged later when light was used to guide reward-seeking behavior)

When did animals learn about the light stimulus as a reward predictor? We noted that the pattern of licking behavior was grossly similar in light and catch trials before the D_2 session, which was the session when catch licks peaked (Figure 1C). Closer examinations revealed that behavioral performances in light and catch trials began to diverge and showed a small but significant difference in the D_2 session (Figure 5A). This suggested that D_2 might be the first session that the light stimulus was used to guide reward-seeking behavior. Such a possibility would predict that BF bursting neurons should begin to respond to the light stimulus in the D_2 session.

We tested this prediction by comparing BF activities between light and catch lick trials in the D_2 session, and found significant differences in all epochs, including the phasic response to the light onset (Figure 5B). This pattern was distinct from the pattern in D_1 sessions, when there was no difference in BF activities in the epochs before exiting the fixation port (Figure 4). We further investigated whether this difference in BF activities between light and catch lick trials emerged within the D_2 session. To test this idea, we compared BF activities at the beginning (first 20 trials) and the end (last 20 trials) of this session (Figure 5C). We found that, at the beginning of the D_2 session, BF activities patterns were similar to the pattern in the D_1 session (Figure 4A), showing no difference between light and catch trials before exiting the fixation port.

However, by the end of the D_2 session, BF responses to the light stimulus had developed in the epochs before exiting the fixation port (Figure 5C). These results support the idea that animals first learned about the light as a reward predictor in the D_2 session.

The divergence between light and catch trials in the D_2 session would correspond to the third step of the retrospective learning process, in which the internal model further expanded and incorporated the light stimulus as a reward predictor (Figure 1E). As a result, the light stimulus acquired the ability to reliably activate BF bursting neurons. After this model expansion, catch lick trials became incompatible with this internal model, and diminished over subsequent sessions.

(Figure 6. Increased BF activity predicted reward-seeking and faster reaction times)

After the D_2 session, we noted that BF phasic responses to light onset continued to grow stronger over sessions (Figure 6A), even though hit rates in light trials already plateaued (Figure 1C). What was the functional significance of BF activity in light trials, and more generally on behavioral performance overall? The results described earlier consistently supported the idea that increased BF activities were associated with increased reward-seeking behaviors (Figures 2-5). Here we provide three additional lines of evidence to further support this idea:

First, in light trials, stronger phasic bursting responses to light onset predicted faster reaction times (RTs) (Figure 6B). As the phasic response to light onset grew stronger over sessions, RTs to light onset decreased accordingly and showed a strong negative correlation at the per session level. **Second**, in light trials where BF responses to the light stimulus had yet to develop (pre-D_2 sessions) and also in catch trials, stronger BF activities after exiting the fixation port predicted reward-seeking behavior (licking) within the same trial type (Figure 6C). In other words, increased BF activities discriminated lick trials from no lick trials within the same trial type. **Third**, increased BF activities before the start of licking in catch lick trials predicted longer durations of licking in individual sessions (Figure 6D). In other words, animals licked for longer durations in the absence of reward when BF activity was higher prior to the start of licking. Together, these observations provided additional support for the idea that the activity of BF bursting neurons promoted reward-seeking behaviors in quantitative ways.

(Figure 7. BF responses to trial outcome were negatively correlated with BF activities in earlier epochs)

Finally, the last question we sought to address was: what kind of information did BF bursting neurons encode, which allowed us to use their activity as a proxy for animals' internal model? We noted that the response of BF bursting neurons to trial outcomes (reward or the absence of reward) often showed the opposite trend as their responses in earlier epochs, both at the per session level (Figure 6A) as well as in single trials (Figure 3B).

We therefore investigated this effect further at the per session level, and found that the amplitude of BF phasic bursting response to the light onset was strongly and negatively correlated with the amplitude of BF responses to the reward in light trials (Figure 7A). Moreover, at the single trial level, we found that the BF evaluation response was negatively correlated with BF responses to the trial outcome, both when the reward was delivered or absent (Figure 7B, 7C).

The presence of strong negative correlations between BF responses to the trial outcome and BF activities in earlier epochs (stimulus onset or evaluation window) is a hallmark feature of prediction error encoding[28]. These observations therefore support the idea that BF bursting neurons encoded reward prediction error in the current behavioral setting, and such information was strongly and quantitatively coupled with reward-seeking behavior.

Discussion

Together, results from the current study support that animals used a retrospective and stepwise learning strategy, which provided a parsimonious framework that can account for the observed behavioral and neurophysiological dynamics (Figure 8). Under this framework, learning was initiated by the unexpected reward at the right reward port (Figure 8, 1st row), which led to the learning of how the reward was predicted by preceding behavioral events in several distinct steps (Figure 8, 2nd-4th row). Behavioral events were sequentially incorporated into the animal's internal model as reward predictors, which started from the event closest to the reward (rightward lick) and sequentially expanded to earlier events (fixation port exit, and then the light stimulus). The manifestation of this stepwise expansion process at the behavioral level was the abrupt emergence of three types of rightward licking behaviors, followed by the sequential elimination of non-rewarded licking behaviors (no-fixation licks and catch licks) (Figures 1C & 3A). At the neuronal level, there were corresponding increases in the activity of BF bursting neurons, which were initially present in all three types of rightward licks (Figures 2E & 3B) and subsequently decreased in non-rewarded licks as those behaviors were sequentially eliminated (Figures 3B & 6A). Moreover, increased BF activities first emerged in the time window closest to the reward (Figures 2E & 3B), while responses to the earlier event (light stimulus) developed later (Figures 4 & 5). Throughout the learning process, the activity of BF bursting neurons encoded reward prediction error signals (Figure 7) and their increased activities consistently predicted reward-seeking behaviors and quantitatively predicted faster reaction times and longer licking durations (Figure 6). These properties enabled us to use BF activities to reconstruct the dynamics of new learning in single trials without perturbing the learning process, and provided an independent validation of the retrospective learning strategy.

(Figure 8. A model depicting the retrospective and stepwise learning strategy and the associated BF neuronal activity)

Under this retrospective learning framework, learning about each behavioral event as a new reward predictor was accompanied by the emergence of BF responses toward that event, which conveyed animals' reward prediction toward that event and promoted the types of reward-seeking behaviors that were compatible with animals' internal model at that step. The stepwise learning strategy therefore offered a parsimonious explaination of why animals committed certain types of 'errors' (non-rewarded licks) at earlier stages of learning, and how such ineffective reward-seeking behaviors were sequentially eliminated in later stages of learning. This highlights the importance of considering all types of reward-seeking behaviors in order to understand the learning process, regardless of whether those behaviors were rewarded or not. The observation of the abrupt behavioral transition in D_1 sessions also suggests that the retrospective learning strategy may underlie the abrupt behavioral transitions observed in other learning contexts[29].

The retrospective and stepwise learning strategy offers three key insights about the learning process. **First**, the same overt behavior can be driven by different internal models at different stages of learning. For example, while the licking response in the new light trial may appear to be the same throughout the learning process, it was in fact driven by different internal models at different stages of learning, which were associated with different activity patterns in the BF (Figure 8). Such covert changes in the internal model underlies a curious observation during early stages of learning (Figure 4): while the light stimulus was clearly-perceptible and had been consistently paired with reward over many trials, the reward-seeking behavior in light trials was not driven by the light stimulus but by other behavioral events (fixation port exit and lick right). The light stimulus had yet to be incorporated into animals' internal reward prediction model at this stage of learning. This observation highlights the possible mismatch between behavioral events in the physical world (such as the light stimulus) and the behavioral events represented in the internal model (the light stimulus was not represented). This is especially a concern during the early stages of learning when the internal reward prediction model may undergo rapid and drastic revisions.

Second, the stepwise expansion of animals' internal models highlights the possibility that the reward-predicting events themselves may change during the learning process. In other words, the learning process not only can incrementally change the weighting of each reward predictor, the reward predictors themselves may also be added or eliminated during the learning process. Such structural revisions of the internal reward prediction model pose a fundamental challenge to theories and models of learning that assume a static set of reward predictors. Using models with incorrect reward predictors will lead to incorrect interpretations, no matter how well the model fits the behavioral performance.

Third, the retrospective and stepwise learning strategy offers an efficient strategy to build the internal reward prediction model: starting from the unexpected reward event and expanding the model by adding one reward predictor at a time. In contrast to the prospective decision tree that contained many state transitions that animals rarely explored (Figure 1D), the retrospective decision tree was much more compact and contained only the relevant subspace that animals did explore (Figures 1E & 8). The compact retrospective model, coupled with the stepwise expansion strategy, likely minimized the cognitive burden of the animal to build the internal reward prediction model during learning.

While our results demonstrated that, at least in the behavioral contexts tested in this study, the retrospective learning strategy was effective and commonly adopted by animals, we do not believe that all associative learning scenarios must involve the retrospective learning strategy. Animals could also employ the prospective learning strategy if, based on past experience, behavioral events could be regarded as potential reward predictors. For example, in the current behavioral context, if animals had previously learned that similar light stimuli predicted reward, the new light stimulus could immediately and prospectively elicit a reward prediction signal and promoted reward-seeking behavior in those trials. The use of prospective strategy, either through generalizations of past experience or resulting from individual variabilities, could facilitate the learning process. Such strategies perhaps underlie the accelerated learning dynamics that we observed in a subset of animal (animal #7, Figure 1C), which also featured the strongest BF responses to the light stimulus during the initial encounter of the new light (Figure 5B).

The results from the current study also highlight the advantages of our approach in understanding the learning dynamics through directly observing the activities of BF bursting neurons. First, BF activities corroborated behavioral observations and provided an independent validation of the retrospective learning strategy. Second, BF activities revealed the covert dynamics of the learning process with single trial temporal resolution. Most importantly, observing BF activities did not perturb the learning process itself. In contrast, many behavioral tests that are commonly used to interrogate the content of learning are invasive in the sense that they inevitably change the learning process that we seek to understand[30,31]. The activities of BF bursting neurons therefore offer a unique opportunity to understand the covert dynamics of the internal model during learning.

Finally, the current study extends our understanding about the roles of BF bursting neurons in the encoding of reward prediction error and in promoting reward-seeking behaviors. Previous studies have demonstrated that BF responses to rewards are negatively modulated by reward expectation[23–26] and support the idea that BF neurons encode a reward prediction error signal[26,27]. The current study further extends this idea and shows that BF bursting neurons also encode reward prediction error in the context of new learning, and such encoding is robust even at the single trial level (Figure 7). This robust encoding of reward prediction error signals therefore quantitatively conveys the amount of reward prediction associated with each behavioral event. The behavioral consequence of this BF reward prediction error signal is the promotion of reward-seeking behaviors. Previous studies have established that BF bursting neurons serve as a bidirectional gain modulation mechanism for reward-seeking behaviors, where increased BF activities promote faster reaction times[23,24] while the inhibition of BF activities leads to rapid behavioral stopping[22]. The current study extends this idea to the context of new learning, by showing that increased BF activities were tightly coupled with reward-seeking behaviors at multiple levels throughout the entire learning process (Figures 2-5), and quantitatively predicted faster reaction times and longer licking durations (Figure 6). Together, these results highlight the functional significance of BF bursting neurons in reward-seeking behaviors. While the neurochemical identity of BF bursting neurons remains unknown, previous studies suggest that BF bursting neurons are likely a subset of GABAergic neurons[20,23,32]. Their neurochemical identity, as well as their specific cellular marker(s), remains to be identified in future studies.

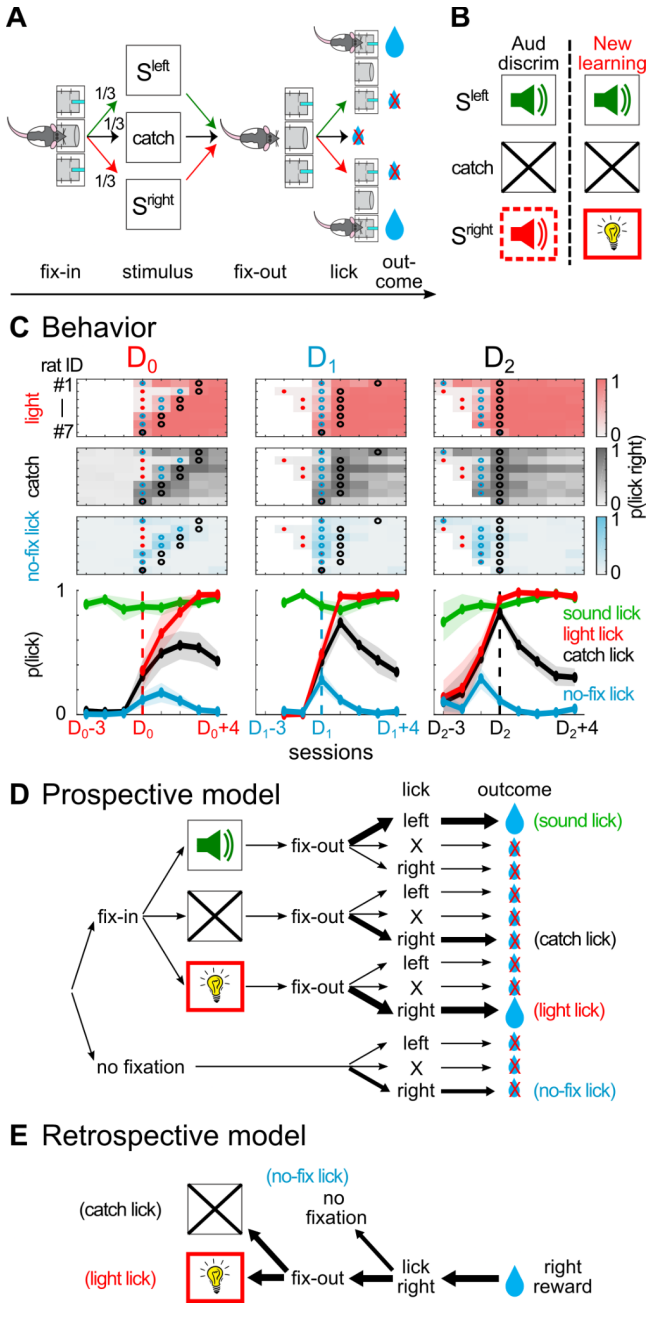


Figure 1. New associative learning led to sequential refinements of reward-seeking behaviors, consistent with a retrospective learning strategy

A, Behavioral task. Rats entered the fixation port to initiate each trial, where they encountered three trial types (S^{left}; S^{right}; catch) with equal probabilities that respectively indicated water reward in the left or right port, or no reward in the case of catch trials (no stimulus). **B**, Rats initially learned an auditory discrimination task (old association phase). At the new learning phase, S^{right} was switched from the sound to a new house light, while other elements of the task remained the same. **C**, The proportion of three types of reward-seeking behaviors toward the right reward port (light licks, catch licks, no-fixation licks) across sessions during new learning (N=7 rats). No fixation licks (cyan) refers to trials in which rats failed to first enter the fixation port before licking in the right reward port. Sessions were respectively aligned, in each column, at the D_0, D_1 or D_2 session of each animal. D_0 refers to the first new learning session with the new light stimulus; D_1 refers to the session when no-fixation licks peaked, which was also when all three types of rightward licks emerged; D_2 refers to the session when catch licks peaked. The D_0, D_1 and D_2 sessions in each animal were indicated by red, cyan, and black circles, respectively. Each row in top panels depicts behavior performance in one animal (#1-7). In the middle and right panels, only sessions in the new learning phase (starting from the D_0 session) were plotted. The emergence, as well as the following sequential elimination, of no-fixation licks and catch licks were observed in all animals. **D**, The decision tree of the new learning phase from a prospective learning strategy, i.e. predicting future states based on the current state. The three types of reward-seeking behaviors toward the right reward port were scattered at different branches of the prospective decision tree. **E**, The decision tree from a retrospective learning strategy, i.e. learning about the states preceding the rewarded state. The three types of reward-seeking behaviors toward the right reward port were closely related in the retrospective decision tree.

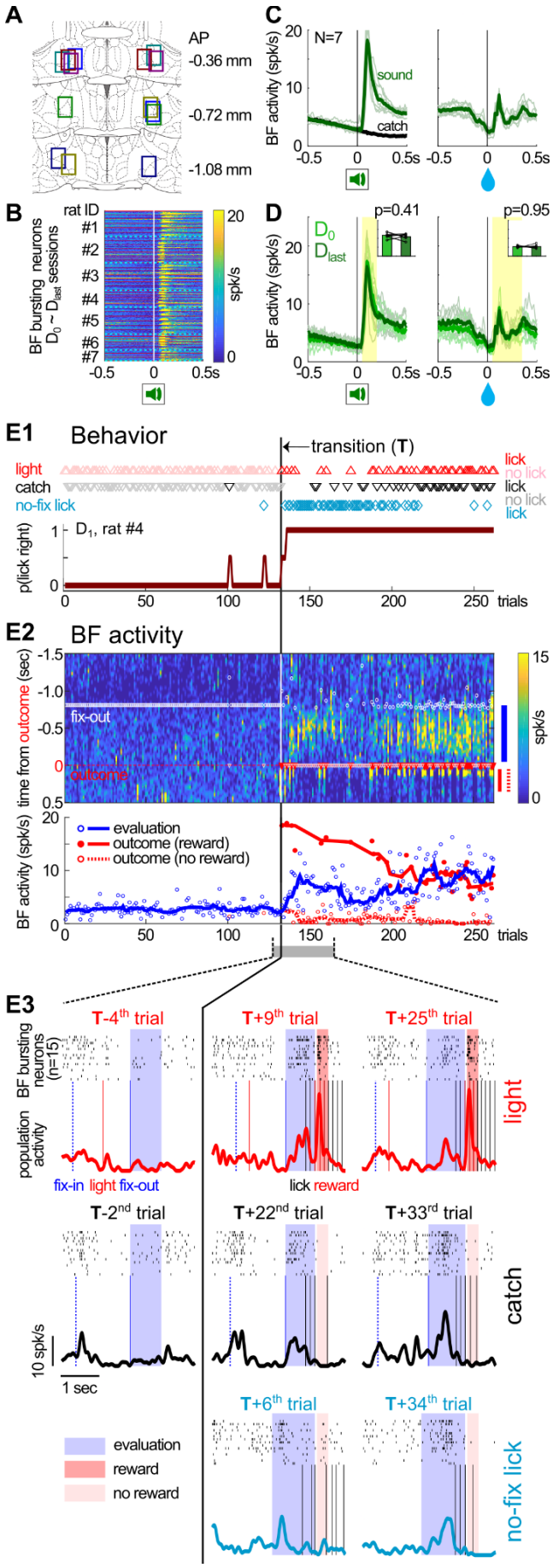


Figure 2. Abrupt transition in reward-seeking behaviors corresponded to increased neuronal activity in the BF during initial learning

A, Locations of electrode bundles targeting bilateral BF (N=7) in coronal sections of the rat brain (coordinates relative to Bregma). Different colors correspond to different animals. **B**, Response of individual BF bursting neurons (n=1013) to the S^{left} sound during new learning sessions (N=45 sessions; separated by thin red lines) in each animal (N=7 rats; separated by cyan dotted lines). BF bursting neurons showed robust and consistent phasic responses to the S^{left} sound throughout the learning process. **C**, Average BF bursting neuron responses to S^{left} sound onset and the associated reward delivery. BF activities in catch trials were plotted for comparison. Responses from individual animals (thin lines) were similar. **D**, The activity of BF bursting neurons remained stable between the first (D_0) and last (D_{last}) recording session. Average activities in the yellow shaded intervals were similar between these two sessions (inset). Thin lines indicate BF activity in individual animals. **E**, Behavioral and BF neuronal dynamics in the D_1 session from a representative animal (rat #4). **E1**, The emergence of three types of rightward licks after the transition point (top), and their combined rightward licking probability across trial types (bottom). The transition point (T) marked an abrupt transition in the pattern reward-seeking behavior that went from no licking to 100% licking. **E2**, Top, population activities of BF bursting neurons (color-coded) in the same trials (X-axis) as shown in E1. Y-axis indicates time in each trial, with time zero aligned at the trial outcome. No lick trials before the transition were aligned instead at the time of fixation port exit (fix-out, white circle) such that the median timing of fix-outs in lick and no lick trials were comparable. The blue and red lines to the right of the panel indicate the time windows for calculating evaluation and outcome responses, respectively. Bottom, BF evaluation responses

and outcome responses across trials. Outcome responses were plotted separately for rewarded (solid red) and non-rewarded (dashed red) licks. Circles indicate BF activities in single trials and lines indicate their respective trends (moving medians). **E3**, Examples of single trial BF activities from the three types of rightward licks taken around the transition point. Each panel showed the spike rasters of BF bursting neurons in this session (n=15) (top), along with the population activity trace and relevant behavioral events (bottom). Shaded intervals indicate the time windows corresponding to evaluation responses (blue) and outcome responses (red) shown in E2. Notice that BF bursting neurons in the same session showed highly similar activity patterns, and that the BF evaluation response rapidly emerged in all three types of rightward licks after the transition point.

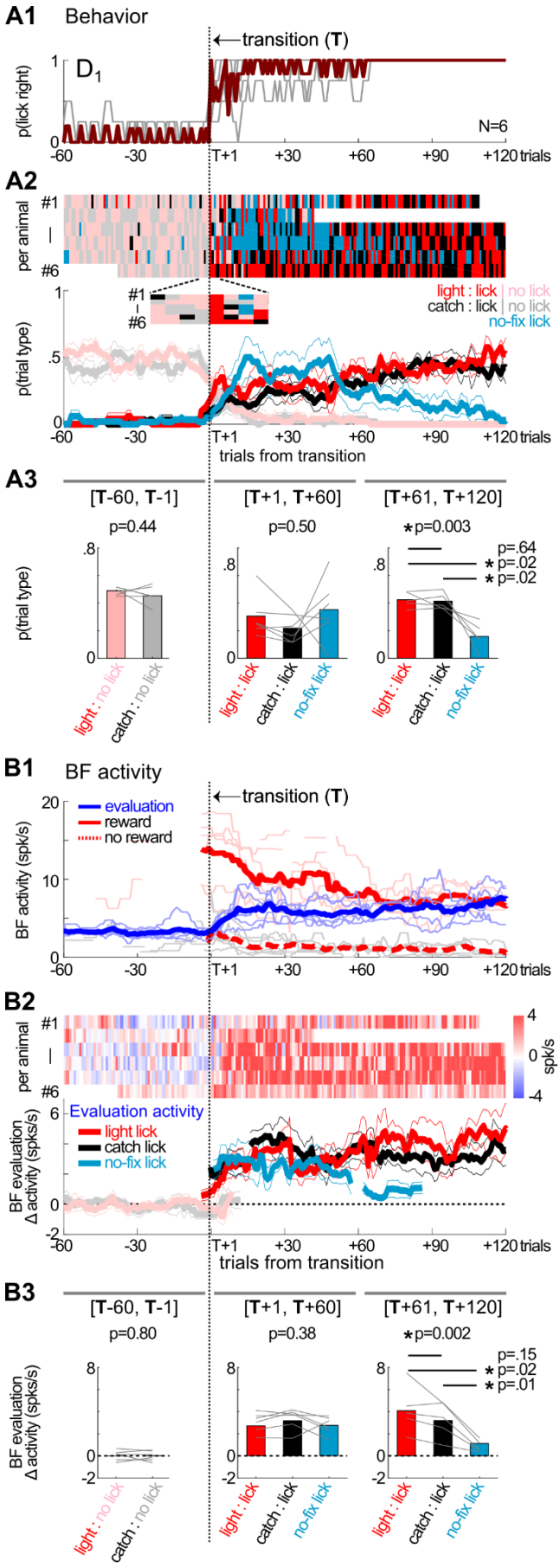


Figure 3. Unexpected reward triggered the rapid emergence and subsequent refinement of reward-seeking behaviors and BF activity

A, The pattern of righward licking behaviors aligned at the transition point in the D_1 session of each animal (N=6). One animal (#7) with accelerated learning dynamics, in which D_1 and D_2 occurred in the same session, was excluded from this analysis. The behavioral response patterns in three trial types (light trials, catch trials and no-fixation licks) were pooled together within each animal (**A1**), or plotted separately (mean ± s.e.m.) (**A2**). The abrupt transition was commonly triggered by the receipt of unexpected reward in the right reward port in light trials (**A2** inset). While all three types of rightward licking emerged immediately after the transition, no-fixation licks subsequently decreased after 60 trials (**A3**). **B**, Corresponding changes in BF activity aligned at the transition point in the D_1 session of each animal (N=6). BF evaluation responses and outcome responses in the three trial types were pooled within each animal (**B1**), as in the example in Figure 2E2. BF evaluation responses were further plotted separately for each trial type, relative to their respective baseline firing rates (mean ± s.e.m.) (**B2**). BF evaluation responses increased similarly in the three types of rightward licking immediately after the transition, and subsequently decreased in no-fixation licks after 60 trials (**B3**). Thin lines in B1 indicate the trend (moving median) of BF activities from individual animals.

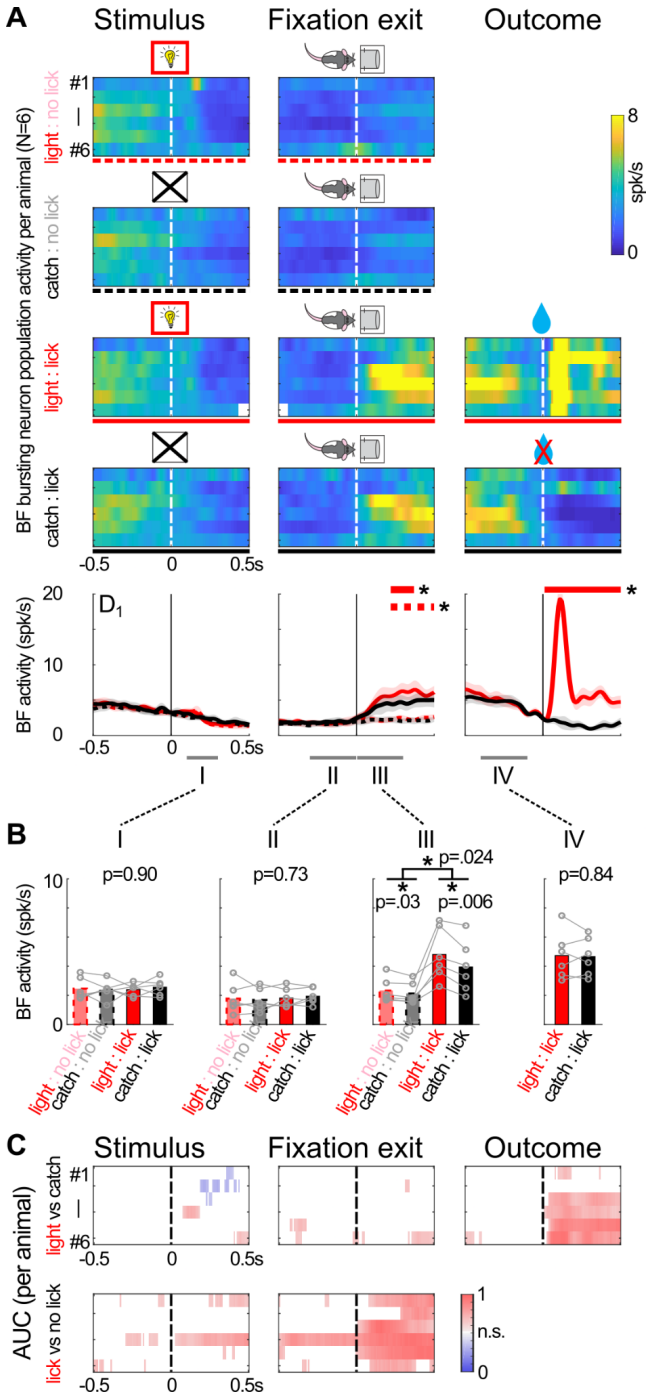


Figure 4. BF neurons did not respond to the new light stimulus during initial learning

A, Population activities of BF bursting neurons in light and catch trials in the D_1 session. BF activities from individual animals (N=6) (top) and group averages (mean ± s.e.m.) (bottom) were plotted separately based on trial type (light or catch) and behavioral response (licking at the right reward port or not), and aligned at three key behavioral events in each trial: stimulus onset, fixation port exit, and trial outcome. Horizontal lines in lower panels indicate significant differences in BF activities ($p<0.01$, 3 consecutive bins) between light lick and catch lick trials (solid red), or between lick and no lick trials (dashed red). **B**, Comparison of BF activities between the four trial combinations during four time windows (I-IV). There was no difference in BF activities before exiting the fixation port (I & II). In contrast, BF activities increased in both light and catch lick trials after exiting from the fixation port (III & IV). BF activities in light lick and catch lick trials were similar before the trial outcome (IV). **C**, Comparison of BF activities between light and catch trials (top) or between lick and no lick trials (bottom) using sliding window ROC analysis, aligned at the three behavioral events. Only significant ($p<0.001$) area-under-curve (AUC) values were shown. BF activities could not reliably discriminate between light and catch trials before the receipt of reward. In contrast, BF activities could reliably discriminate between lick and no lick trials after exiting the fixation port.

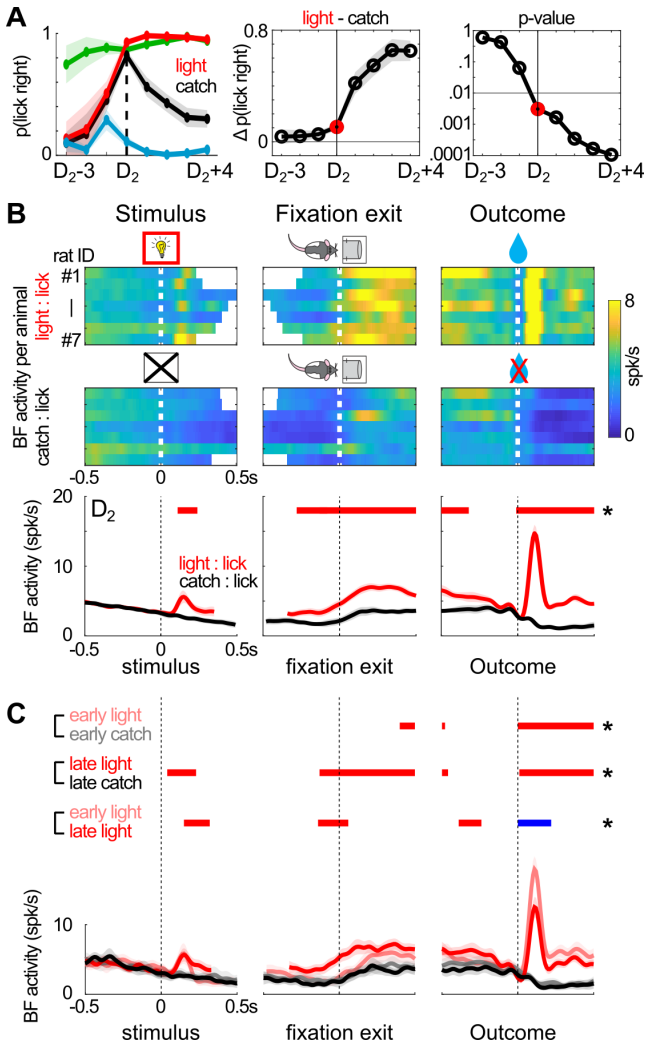


Figure 5. BF responses to light onset emerged later when light was used to guide reward-seeking behavior

A, Behavioral performance aligned to the D_2 session of each animal (left) (N=7), the same as the right panel in Figure 1C. The probability of licking toward the right reward port in light and catch trials began to differ significantly at the D_2 session (middle; right). **B**, Population activities of BF bursting neurons in light lick and catch lick trials in the D_2 session for individual animals (top) (N=7) and their group average (mean ± s.e.m.) (bottom). BF activities in light lick trials were significantly higher in epochs before exiting the fixation port. Significant differences in BF activities ($p<0.01$, 3 consecutive bins) were indicated by horizontal red lines. BF activities were truncated at the median RT of the respective sessions (see Methods for details). **C**, BF activities (mean ± s.e.m.) in light lick and catch lick trials at the first 20 trials (early) or the last 20 trials (late) of the D_2 session. At the beginning of the D_2 session, BF activities between light lick and catch lick trials were similar in the interval between stimulus onset and fixation port exit. Differences in BF activities in this interval became significant by the end of the D_2 session. Similar differences in this interval were observed when comparing BF activities between the early and late light lick trials. Significant excitation (red) or inhibition (blue) in BF activities ($p<0.01$, 3 consecutive bins) were indicated by horizontal lines.

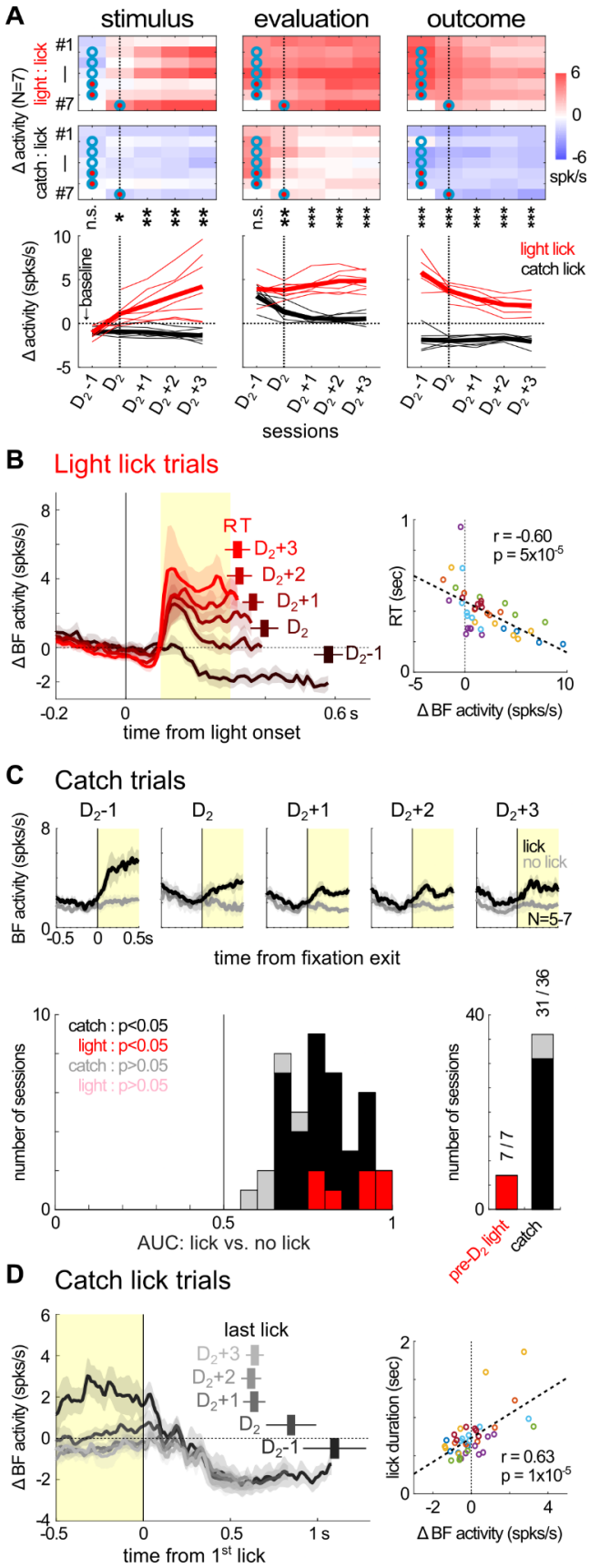


Figure 6. Increased BF activity predicted reward-seeking and faster reaction times

A, The evolution of BF activities in light and catch lick trials, aligned at each animal's D_2 session, calculated relative to their respective baseline firing rates. Significant differences in BF activities between the two trial types were indicated (*p<0.05; **p<0.01; ***p<0.001). BF activities in the stimulus onset and evaluation epochs diverged in the D_2 session and their differences grew larger afterwards. Also notice that BF responses to the reward in light lick trials decreased over sessions. **B**, Average responses of BF bursting neurons (mean ± s.e.m.) to light onset in light lick trials (left). BF activities, relative to their respective baseline firing rates, were truncated at the median RT of the respective sessions. The RTs (mean ± s.e.m.) of the corresponding sessions were shown above each trace. Stronger BF responses to light onset in the [0.1, 0.3]s window (yellow shaded interval, left panel) were correlated with faster RTs in individual sessions (right). Each circle indicates one session and different colors correspond to different animals. **C**, Average responses of BF bursting neurons (mean ± s.e.m.) in catch trials aligned at fixation port exit, plotted separately for lick and no lick trials (top). Distributions of AUC values from comparing BF activities in the [0, 0.5]s window after fixation port exit (yellow shaded interval) between lick and no lick trials within the same trial type (catch or light trials) (bottom). Within the same trial type, increased BF activities reliably predicted reward-seeking behavior toward the right reward port. Only light trials from pre-D_2 sessions were included because BF responses to the light stimulus had not developed. **D**, Average responses of BF bursting neurons (mean ± s.e.m.) in catch lick trials aligned at the first lick (left). BF activities, relative to their respective baseline firing rates, were truncated at the median lick duration of the respective sessions. The timing of the last lick (mean ± s.e.m.) of the corresponding sessions

were shown above each trace. BF activities prior to the start of licking in catch lick trials (yellow shaded interval, left panel) were positively correlated with the median lick duration in individual sessions (right). Each circle indicates one session and different colors correspond to different animals.

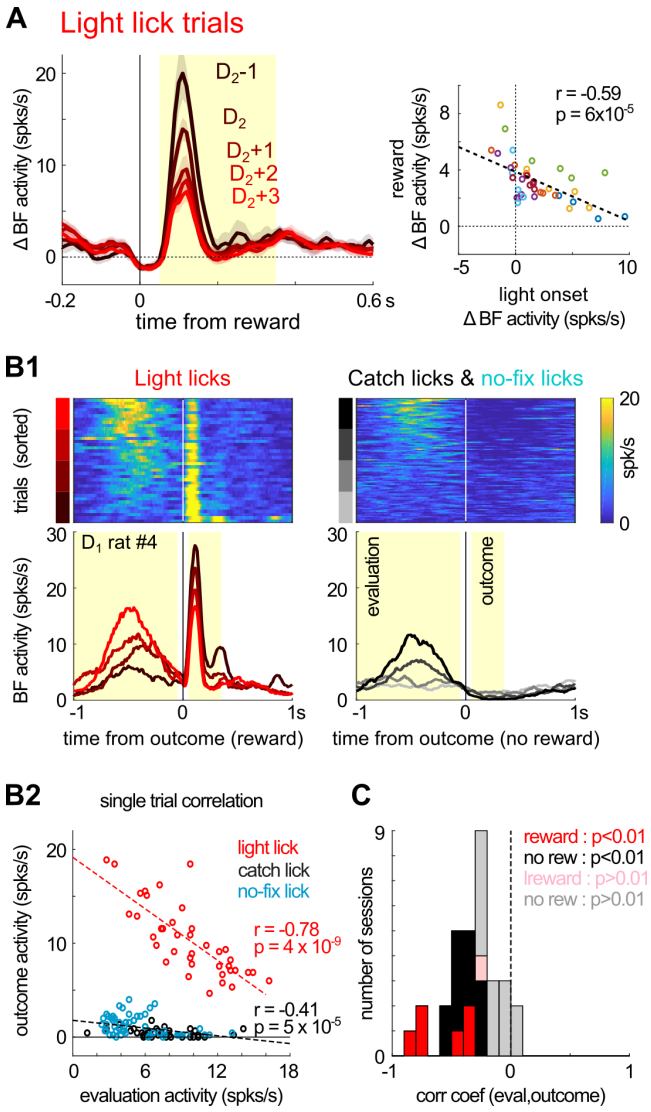


Figure 7. BF responses to trial outcome were negatively correlated with BF activities in earlier epochs

A, Average responses of BF bursting neurons (mean ± s.e.m.) to the reward in light lick trials, relative to their respective baseline firing rates, plotted separately for the five sessions relative to the D_2 session (left). BF responses to the reward (yellow shaded interval, left panel) were negatively correlated with BF responses to light onset (Figure 6B) in individual sessions (right). Each circle indicates one session and different colors correspond to different animals. **B**, Negative correlation of BF activities between evaluation and outcome responses in single trials. **B1**, Single trial BF activities in an example D_1 session, aligned at the trial outcome. Single trial BF responses were plotted separately for rewarded licks (light licks) and non-rewarded licks (catch licks and no-fixation licks). Yellow shaded intervals indicate time windows for calculating evaluation and outcome responses. Trials were sorted by the amplitude of evaluation responses (top). Average BF activities from the four quartiles of trials were plotted separately (bottom). **B2**, Negative correlation between single trial BF evaluation responses and outcome responses in this session. Each dot represents one trial. **C**, Histogram of correlation coefficients between evaluation and outcome responses from individual sessions. Results for rewarded lick trials (light licks) were calculated from pre-D_2 sessions when BF responses to the light stimulus had not developed (N=7), and for non-rewarded lick trials from sessions with at least 50 trials of catch and no-fixation licks combined (N=25). Most sessions showed significant negative correlations.

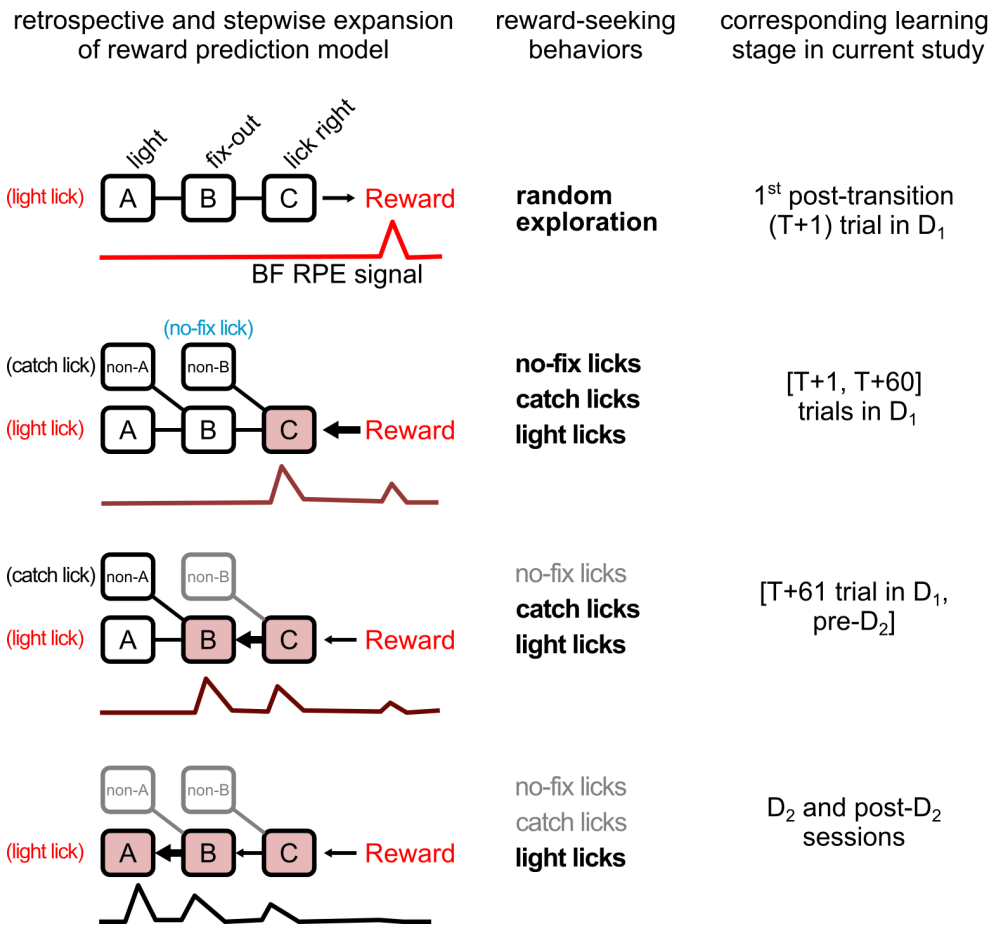


Figure 8. A model depicting the retrospective and stepwise learning strategy and the associated BF neuronal activity

The behavioral and neural dynamics observed in this study support a retrospective and stepwise learning model. In our task, new learning requires animals to learn that the reward in the right port is associated with a sequence of behavioral events A-B-C (light onset - exit fixation - lick at the right reward port). The first reward during the new learning was obtained through random exploration when animals accidentally completed this sequence, with the unexpected reward eliciting a strong BF bursting response (1st row). This triggered a retrospective learning process to learn about the state preceding the unexpected reward (i.e., event C, 2nd row). This process was accompanied by the emergence of BF responses toward the newly incorporated event, as well as reduced BF responses to the reward. Subsequently, this retrospective learning process underwent stepwise expansions to sequentially incorporate earlier events (B, and then A; 3rd and 4th rows). At each step of learning, BF activities reflected the reward prediction error signals associated with each event and promoted reward-seeking behaviors containing those learned events (filled red boxes). This model explains how non-rewarded licking behaviors were sequentially eliminated (gray). As a result, reward-seeking behaviors became more selective in later stages of learning as the model incorporated more reward-predicting behavioral events. The corresponding behavioral landmarks for each learning step are indicated in the right column.

Animal ID	S^{left} sound	S^{right} sound (auditory discrimination)	S^{right} light (new learning)
#1	12kHz 80dB 2s	6kHz 80dB 2s	Center Light 0.5s
#2	100Hz clicker 75dB 1s	white noise 75dB 1s	Center Light 1s
#3	100Hz clicker 75dB 1s	white noise 75dB 1s	Center Light 1s
#4	100Hz clicker 75dB 1s	white noise 75dB 1s	Center Light 1s
#5	100Hz clicker 75dB 1s	white noise 75dB 1s	Center Light 1s
#6	100Hz clicker 75dB 1s	white noise 75dB 1s	Center Light 1s
#7	100Hz clicker 75dB 1s	white noise 75dB 1s	Center Light 1s

Table 1. Stimulus parameters of the S^{left} and S^{right} stimuli for each animal

	Sessions (relative to each animal's D_2 session)							
Animal ID	D_2 -3	D_2 -2	D_2 -1	D_2	D_2 +1	D_2 +2	D_2 +3	D_2 +4
#1	4/13	1/4	23/32	25/33	34/49	28/41	31/46	29/42
#2	26/43	29/46	26/38	25/40	25/42	13/28	15/31	18/36
#3		33/39	20/29	29/38	26/36	21/33	26/34	30/40
#4		17/27	15/21	17/23	21/30	21/25	24/26	
#5			29/35	32/39	30/38	30/39	29/36	31/35
#6			17/30	16/29	20/37	16/28	18/29	20/27
#7				23/28	20/22	11/13	19/23	

bursting neurons / # BF neurons

D_0 D_1 D_2 session symbol

Table 2. Timing of the three landmark sessions in each animal and the number of recorded BF neurons in each session

Methods

Ethics Statement

All experimental procedures were conducted in accordance with the National Institutes of Health (NIH) Guide for Care and Use of Laboratory Animals and approved by the National Institute on Aging Animal Care and Use Committee.

Subjects

Seven male Long Evans rats (Charles River, NC), aged 3–6 months and weighing 300–400 grams were used for this experiment. Rats were housed in a 12/12 day/night cycle and were provided with 10 to 12 dry pellets per day and unrestricted access to water. Rats were first trained in daily sessions lasting 60-90 minutes.

Apparatus

Twelve plexiglass operant chambers (11” L×8 ¼” W×13” H), custom-built by Med Associates Inc. (St. Albans, VT), were contained in sound-attenuating cubicles (ENV-018MD) each with an exhaust fan that helped mask external noise. Each chamber’s front panel was equipped with an illuminated nose poke port (ENV-114M) located in the center (horizontal axis) as the fixation port, which was equipped with an infrared (IR) sensor to detect the entry of the animals’ snout into the port. On each side of the center nose-poke port there were two reward ports (CT-ENV-251L-P). Two IR sensors were positioned to detect reward-port entry and sipper-tube licking, respectively.

Sucrose solution (13.3%) was used as reward and delivered through the sipper tubes located in the reward ports. Reward delivery was controlled by solenoid valves (Parker Hannifin Corp #003-0111-900, Hollis, NH) and calibrated to provide 10 µl of solution per drop. Each chamber was equipped with a ceiling-mounted speaker (ENV-224BM) to deliver auditory stimuli, and a stimulus light (ENV-221) positioned above the center fixation port to serve as the new light stimulus. Behavioral training protocols were controlled by Med-PC software (Version IV, Med Associates Inc.), which stored all event timestamps at 2 ms resolution and also sent out TTL signals to the neurophysiology recording systems.

Behavioral Training Procedures

Food-restricted adult Long-Evans rats were trained in operant chambers that were dimly lit. Rats were first trained in an auditory discrimination task. After an inter-trial interval (4-6 s), each trial started with the illumination of the center fixation port, which was turned off when rats poked the fixation port. Rats were required to maintain fixation in the center nose poke port for a variable amount of foreperiod. Four different foreperiods (0.35, 0.5, 0.65, and 0.8 s) were used, pseudorandomly across trials, to minimize temporal expectation of stimulus onset. After the foreperiod, one of three conditions was randomly presented with equal probabilities: a S^{right} sound stimulus indicated reward on the right port; a different S^{left} sound stimulus indicated reward on the left port; or the absence of stimulus (catch trial) indicated no reward. The specific sound stimuli chosen as the S^{right} and S^{left} for each animal are listed in TABLE 1. An internal timestamp was recorded in catch trials to mark the onset of the would-be stimulus. Early fixation port exit before the foreperiod or licking in the incorrect side port resulted in no reward delivery

and reset the timer for the inter-trial-interval. Licking in the correct port within a 3 s window led to three drops of sucrose water reward, delivered starting at the 3rd lick. The delivery of reward at the 3rd lick created an expectation for trial outcome at this time point, which was dissociated in time from the initiation of licking (1st lick). We will therefore refer to the time point of the 3rd lick also as the trial outcome event.

After reaching asymptotic behavioral performance in the auditory discrimination task, rats underwent stereotaxic surgery to implant moveable electrode bundles into bilateral BF. After surgery, rats were re-trained in the auditory discrimination task to asymptotic performance level. Once stable recording of BF neuronal activity was achieved, a new learning phase was introduced by replacing the S^{right} stimulus from the sound used in the auditory discrimination task by the central light above the fixation port to indicate reward on the right port. All other aspects of the task remained the same. See TABLE 1 for details of the S^{right} and S^{left} stimuli for each animal.

Stereotaxic Surgery and Electrode

Surgery was performed under isoflurane anesthesia as previously described[24]. Multiple skull screws were inserted to anchor the implant, with one screw over the cerebellum serving as the common electrical reference and a separate screw over the opposite cerebellum hemisphere serving as the electrical ground. Craniotomies were opened to target bilateral BF (AP –0.6 mm, ML ±2.25 mm relative to Bregma)[33]. The electrode contained two bundles of 16 polyimide-insulated tungsten wires (38 μm diameter; California Fine Wire, CA), each bundle ensheathed in a 28-gauge stainless steel cannula and controlled by a precision microdrive. The impedance of individual wire was ~ 0.1 MΩ measured at 1 kHz (niPOD, NeuroNexusTech, MI). During surgery, the cannulae were lowered to DV 6.5 mm below cortical surface using a micropositioner (Model 2662, David Kopf Instrument) at a speed of 50 μm/s. After reaching target depth, the electrode and screws were covered with dental cement (Hygenic Denture Resin), and electrodes further advanced to 7.5 mm below the cortical surface. Rats received ibuprofen and topical antibiotics after surgery for pain relief and prevention of infection, and were allowed one week to recover with *ad libitum* food and water. Cannulae and electrode tip locations were verified with cresyl violet staining of histological sections at the end of the experiment. All electrodes were found at expected positions between AP [–0.2, –1.2] mm, ML [1.5, 3] mm, relative to Bregma, and DV [7.5, 8.5] mm relative to cortical surface (Figure 2A).

Data acquisition and spike sorting

Electrical signals were referenced to a common skull screw placed over the cerebellum. Electrical signals were filtered (0.3 Hz–7.5 kHz) and amplified using Cereplex M digital headstages and recorded using a Neural Signal Processor (Blackrock Microsystems, UT). Single unit activity was further filtered (250 Hz–5 kHz) and recorded at 30 kHz. Spike waveforms were sorted offline to identify single units using custom software written in Python. Only single units with clear separation from the noise cluster and with minimal (<0.1%) spike collisions (spikes with less than 1.5 ms interspike interval) were used for further analyses, consistent with previous studies of BF bursting neurons[19–24]. Additional cross-correlation analysis was used to remove duplicate units recorded simultaneously across multiple electrodes[19–24].

Recording during the new learning phase

After surgery, BF neuronal activity was monitored while rats were re-trained in the auditory discrimination task to asymptotic performance level. During this re-training phase, BF electrode depths were adjusted slightly (by advancing electrodes at 125 μm increment) until a stable population of BF single units can be recorded. At this point, the new learning phase with the light as the new S^{right} stimulus was introduced and rats were trained and recorded daily with BF electrodes remained at the same depth. This approach allowed us to monitor the activity of a large population of BF neurons and follow its temporal evolution across sessions.

Data analysis

Data were analyzed using custom Matlab (MATLAB The MathWorks Inc., Natick, MA) scripts.

Define different behavioral response types

Licking responses were defined for stimulus (S^{left} and S^{right}) and catch trials if rats licked at least three times in the reward port within the 3 sec window after stimulus onset (or the corresponding timestamp for the would-be stimulus in catch trials). Licking responses to the correct reward port were rewarded with three drops of sucrose water, delivered starting at the 3rd lick (referred to as trial outcome event). During the new learning phase, licking responses in light and catch trials were predominantly to the right reward port. Responses to the left reward port were negligible and therefore not analyzed.

No-fixation licks corresponded to licking responses to the right reward port that were not preceded by poking the center fixation port. Specifically, no-fixation licks were defined based on three criteria: (1) rats made at least three consecutive licks in the right reward port; (2) the interval between the last exit from the fixation port and the first lick must be greater than 2 sec; (3) the interval between the last exit from the reward port and the subsequent first lick in the same reward port must be greater than 1 sec. These duration thresholds were determined based on the empirical licking patterns across animals. No-fixation licks to the left reward port occurred at much lower frequencies and therefore not analyzed. In the analyses of learning dynamics in the D_1 session (Figures 2 & 3), no-fixation licks were treated as rightward licking trials, even though such behaviors were self-initiated and not imposed by the task design.

Reaction time (RT) in light trials in a session (Figure 6B) was defined as the median of the interval between the onset of the light stimulus and the exit from the fixation port in light lick trials. Lick duration in catch trials in a session (Figure 6D) was defined as the median of the interval between the first and the last lick in catch lick trials.

Define the D_0, D_1 and D_2 learning landmarks

During the new learning phase, three sessions (D_0, D_1, D_2) were identified in individual animals as landmarks that demarcated distinct stages of new learning (Figure 1C). The D_0 session was defined as the very first session the new light stimulus was introduced. The D_1 session was defined as the first session when animals began to respond correctly in light trials, which also corresponded to the session in which no-fixation licks occurred most frequently. The D_2 session was defined as the session in which catch licks occurred most frequently. The D_1 and D_2 landmarks allowed us to identify similar learning stages across animals despite their individual differences in learning dynamics. The specific timing of the three landmark sessions in each animal are provided in TABLE 2 (also see Figure 1C). One animal (ID#7) with accelerated

learning dynamics, in which D_1 and D_2 occurred in the same session, was excluded from analyses of D_1 neural dynamics (Figure 3-4) to ensure that neural activities associated with D_2 did not confound the neural dynamics in the D_1 session. The BF neuronal activity in this animal was included in the analysis of D_2 neural dynamics (Figure 5) and showed the strongest phasic response to the light onset among all animals, consistent with its accelerated learning dynamics.

Identification of the behavioral transition point in the D_1 session

The behavioral transition points in D_1 sessions (Figures 2 & 3) were identified based on behavioral response patterns in three trial types combined: light trials, catch trials and no-fixation licks. The behavioral response pattern in each trial was coded as either 1 or 0 based on whether animals licked in the right reward port in that trial. The behavioral transition point was defined as the point with the largest difference in licking responses between the 20 trials before and the 20 trials after that point. In 5/7 animals, the first trial after the behavioral transition was a rewarded light lick trial. In the other two animals, the transition point was adjusted to the closest light lick trial by 2 or 4 trials, respectively.

Identification of BF bursting neurons

BF bursting neurons were defined as BF single units whose average firing rates during the [0.05, 0.2]s window after stimulus onset increased by more than 2 spikes/s in the S^{left} sound trials compared to the corresponding window in catch trials (Figure S1). This contrast between sound trials and catch trials was necessary because many BF neurons changed their activity during the foreperiod while waiting for stimulus onset. This subtraction procedure removed the nonstationary baseline before stimulus onset and allowed us to ask whether BF neurons truly responded to the sound stimulus. In addition, BF bursting neurons should have baseline firing rates (during the [-1, 0]s window relative to the trial start signal) less than 10 spikes/s.

A total of 1453 BF single units were recorded over 45 sessions (N=7 rats), of which 70% (1013/1453) were classified as BF bursting neurons based on their stereotypical phasic response to the S^{left} sound (22.5±7.3 neurons per session, mean±std) (Figure S1, TABLE 2). One session with only one BF bursting neuron was excluded from the analysis of BF population activities. The large number of BF bursting neurons recorded in each session allowed us to treat them as a representative sample of all BF bursting neurons, whose responses to the S^{left} sound were highly stable throughout the learning process (Figure 2B-D). This strategy ensured that we were following functionally the same neuronal ensemble and could track how BF bursting neurons acquired responses to the new light during learning, regardless of whether the identities of these BF neurons were exactly the same in each session.

Population BF responses to behavioral events

The spike timestamps of all BF bursting neurons in a single session were pooled together to approximate the population activity of all BF bursting neurons. Population peri-stimulus time histograms (PSTHs) were calculated with 10-ms bins, and normalized by the number of BF bursting neurons in a session.

To properly assess whether BF bursting neurons responded to the onset of the new light stimulus (Figures 4-6), it was important to disambiguate such stimulus-onset responses from

the increased BF activities after fixation port exit (Figure 4). To achieve this goal, PSTHs to the stimulus onset were calculated based only on spikes that occurred before fixation port exit in individual trials, resulting in different interval lengths (between stimulus onset to fixation port exit) across trials. Accordingly, the calculation of the mean PSTH across all trials in a session was adjusted for the different number of trials at different interval lengths. The mean PSTHs were further truncated at the median interval length of that session to reduce noisy estimates of PSTHs at long interval lengths due to lower number of trials. When PSTHs were averaged across animals, the averaged PSTHs were further truncated at the mean of median interval latencies across animals. This truncation procedure resulted in the uneven lengths of PSTHs across individual animals (Figures 4 & 5) and across sessions (Figure 6). This procedure was also applied to calculating the BF responses before fixation port exit to include only spikes that occurred after stimulus onset (Figures 4 & 5), and for calculating BF responses during licking in catch lick trials (Figure 6D).

The time windows used to quantify average BF activity in different epochs were indicated in respective figures, and corresponded to the following: [0.05, 0.2]s after S^{left} sound onset; [0.1, 0.3]s after S^{right} light stimulus onset; [0.1, 0.3]s after the timestamp for the would-be stimulus in catch trials; [0.05, 0.35]s after the 3rd lick for outcome responses; [-0.3, 0]s and [0, 0.3]s relative to the fixation port exit. The epoch for calculating evaluation response is described below.

Evaluation response (Figures 2E, 3B, 7B, 7C) refers to the increased BF activity after exiting the fixation port and before the trial outcome (3rd lick). The evaluation response reflected animals' internal evaluation because no additional sensory stimuli were presented during this epoch. Specifically, the evaluation response was calculated in individual trials and defined as the maximum firing rate of any 500ms window during the evaluation epoch, which corresponded to the interval between [fix-out, outcome], with additional adjustments according to trial types. In light lick and catch lick trials, the evaluation epoch was defined as [fix-out, outcome] in each trial. The epoch durations in light lick and catch lick trials within each session were used as the reference point for other trial types as described next. In no-fixation licks, in which the fix-out event was absent, the duration of evaluation epoch was set as the 95th percentile of the evaluation epoch durations in light lick and catch lick trials. In light no lick and catch no lick trials, in which the 3rd lick event was absent, the duration of the evaluation epoch was set as the median of the evaluation epoch durations in light lick and catch lick trials. These adjustments in the definition of evaluation epochs, as well as its calculation of maximum firing rate within the epoch, took into consideration the behavioral variability across trial types, learning stages and individual animals.

To evaluate the dynamic changes of BF activities around the transition point in the D_1 session (Figures 2E & 3B), single trial evaluation and outcome responses were smoothed using moving median over 10 trials. The smoothed trends were aligned at the transition point and then averaged across all animals. Only trials with smoothed trend data from at least 4 animals were plotted in the group average (Figure 3B).

Statistics

Statistical comparisons were conducted using the Statistics and Machine Learning Toolbox (version 11.3) in MATLAB (R2018a) (https://www.mathworks.com/). Paired t-test (ttest.m) was used to compare behavioral and neural activity differences between two groups (Figures 2D,

3A3, 3B3, 5A, 6A). Repeated measures analysis of variance (ranova.m) was used for comparisons involving more than 2 groups, by specifying the appropriate within-subject models (Figures 3A3, 3B3, 4B). Comparisons of PSTHs between two groups (Figures 4A, 5B, 5C) was conducted for each 100 ms sliding window (10 ms step) using paired t-test. Significance level was set at $p<0.01$ for three consecutive bins. Pearson correlation (corrcoef.m) was used to determine the relationship between neuronal activities and/or behavior (Figures 6B, 6D, 7).

Receiver operating characteristic (ROC) and area under curve (AUC) analysis

To determine whether the activity of BF bursting neurons differentiated between trial types within each D_1 session (Figure 4C), we compared BF activity for each 100ms sliding window (10ms step) using the AUC measure of ROC analysis (auc.m by Alois Schloegl). At each sliding window, BF population activity was calculated for each light and catch trial, and distributions of BF activities were compared between light vs catch trials or between lick vs no lick trials. Significance level was set at $p<0.001$ using 10,000 trial-shuffled random permutations.

To determine whether BF activity differentiated between lick and no lick trials within the same trial type (light or catch trials) (Figure 6C), we compared BF activity in the [0, 500] ms window after exiting the fixation port. For each session and each trial type, lick and no lick trials must each constitute at least 10% of that trial type to be included in the analysis. Catch trials from all sessions were included in this analysis. Only light trials before the D_2 session (pre-D_2) were included in this analysis because BF responses to the onset of the light stimulus had not developed in those sessions. Significance level was set at $p<0.05$ using 1,000 trial-shuffled random permutations.

Acknowledgements

We thank P.R. Rapp, A. Scaglione, S.W. Wu and K.L. Hsieh for critical discussions of the manuscript; B.M. Brock for technical support. This research was funded by the Intramural Research Program of the National Institute on Aging (NIH, USA) and by NARSAD Young Investigator Award to S.L.. Additional supports came from the Ministry of Science and Technology (MOST, Taiwan) grants (107-2320-B-010-028-MY3; 108-2638-B-010-002-MY2; 110-2628-B-A49A-505), and from the Brain Research Center, National Yang Ming Chiao Tung University from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

Author Contributions

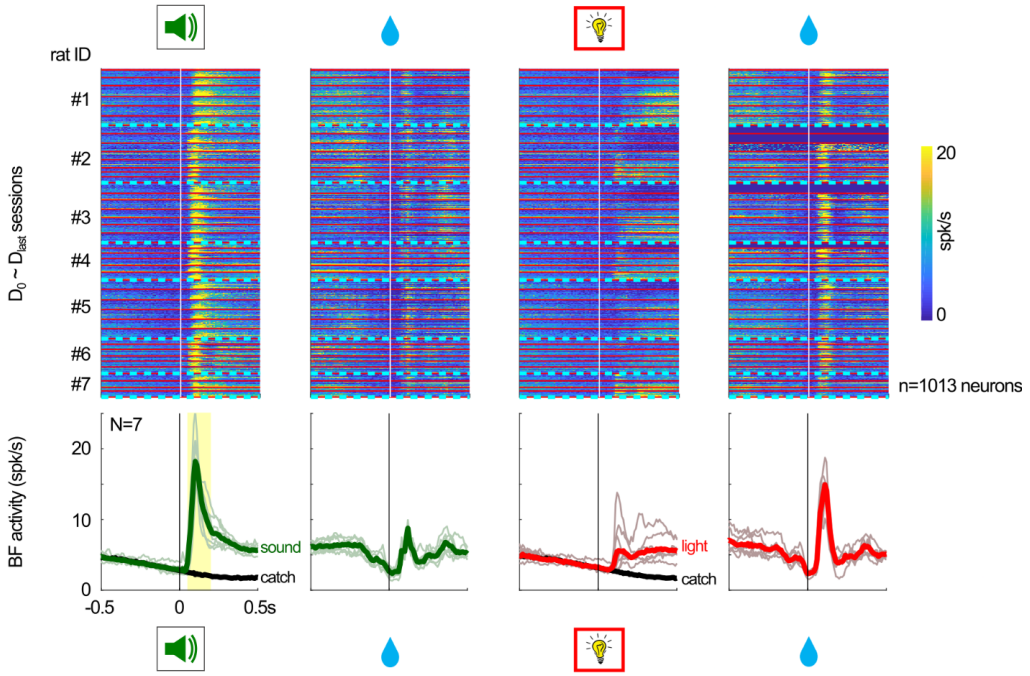
H.E.M. and S.L. designed the study. H.E.M. and K.V. performed experiments and collected data. H.E.M. and S.L. analyzed data. H.E.M. and S.L. wrote the manuscript with inputs from K.V.

Author Information

The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.L. (shihchieh.lin@nycu.edu.tw).

Supplementary Figure

BF bursting neurons



Other BF neurons

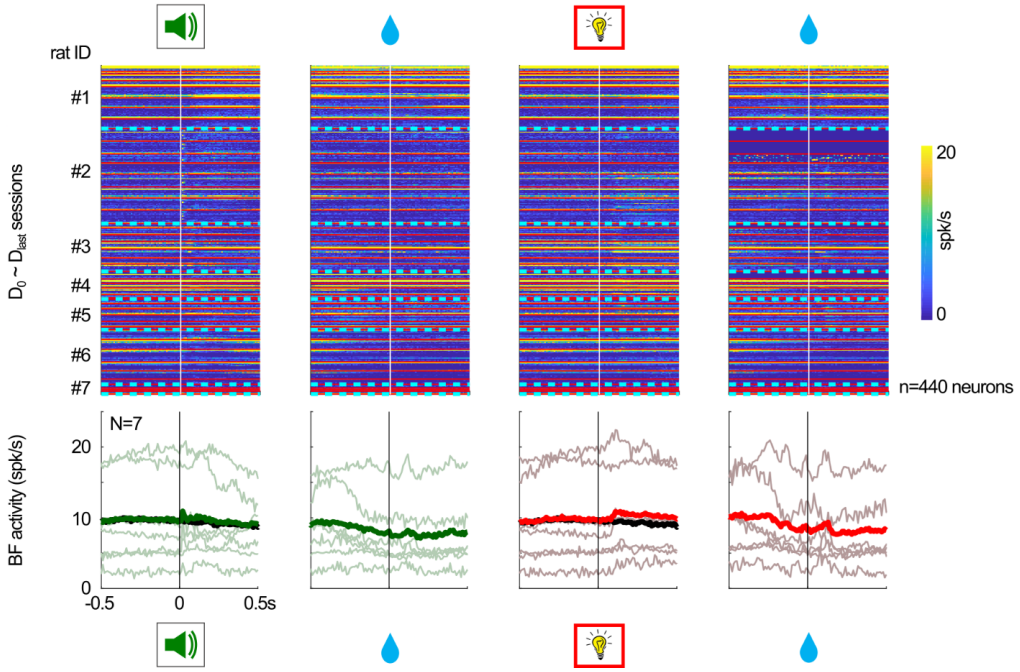


Figure S1. Identification of BF bursting neurons and their responses to behavioral events

Responses of individual BF bursting neurons (n=1013) and other BF neurons (n=440) to behavioral events (stimulus onset and reward) in the S^{left} sound trials (two left columns) and in the S^{right} light trials (two right columns). Recordings from different sessions (N=45 sessions; separated by thin red lines) and different animals (N=7 rats; separated by cyan dotted lines) were separated by horizontal lines. Lower panels showed the average response (thick lines) pooled across individual animals (thin lines). BF activities in catch trials (thick black lines) were plotted for comparison. Responses to the stimulus onset event were calculated based on all trials in that session, regardless of subsequent behavioral responses (licking or not). On the other hand, responses to the reward were calculated based only on correct licking trials. Conventions as in Figure 2B-C.

BF bursting neurons were defined as BF single units whose average firing rates during the [0.05, 0.2]s window after stimulus onset (yellow shaded interval) increased by more than 2 spikes/s in the S^{left} sound trials compared to the corresponding window in catch trials. This contrast between sound trials and catch trials was necessary because many BF neurons changed their activity during the foreperiod while waiting for stimulus onset. In addition, BF bursting neurons should have baseline firing rates less than 10 spikes/s. The activities of BF bursting neurons in S^{left} sound trials were highly similar across sessions and across animals (Figure 2C-D).

Note that the calculation of these PSTHs to the stimulus onset event (as well as those in Figure 2B-D) did not exclude spikes that occurred after fixation port exit (as in Figures 4-6). The truncation procedure used in Figures 4-6 was needed to disambiguate BF responses to light onset from the increased BF activities after fixation port exit (i.e. evaluation responses).

References

1. Pavlov, I. P. *Lectures on conditioned reflexes: Twenty-five years of objective study of the higher nervous activity (behaviour) of animals*. vol. 414 (Liverwright Publishing Corporation, 1928).
2. Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. in *Classical conditioning: current research and theory* vol. 2 (1972).
3. Thorndike, E. L. Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements* **2**, i (1898).
4. Dayan, P. & Balleine, B. W. Reward, motivation, and reinforcement learning. *Neuron* **36**, 285–298 (2002).
5. O'Doherty, J. P., Hampton, A. & Kim, H. Model-based fMRI and its application to reward learning and decision making. in *Annals of the New York Academy of Sciences* vol. 1104 35–53 (2007).
6. Doll, B. B., Simon, D. A. & Daw, N. D. The ubiquity of model-based reinforcement learning. *Curr. Opin. Neurobiol.* **22**, 1075–1081 (2012).
7. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189–208 (1948).
8. Wilson, R. C., Takahashi, Y. K., Schoenbaum, G. & Niv, Y. Orbitofrontal cortex as a cognitive map of task space. *Neuron* **81**, 267–279 (2014).
9. Behrens, T. E. J. *et al.* What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron* **100**, 490–509 (2018).
10. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
11. Eshel, N. *et al.* Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* **525**, 243–246 (2015).
12. Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J. & Frith, C. D. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* **442**, 1042–1045 (2006).

13. Sutton, R. S. & Barto, A. G. Reinforcement Learning: An Introduction. *IEEE Trans. Neural Netw.* **9**, 1054–1054 (1998).
14. Russell, S. J. & Norvig, P. *Artificial Intelligence: A Modern Approach*. 1152 (Prentice Hall, 2010).
15. K Namboodiri, V. M. & Stuber, G. D. The learning of prospective and retrospective cognitive maps within neural circuits. *Neuron* **109**, 3552–3575 (2021).
16. Gallistel, C. R., Craig, A. R. & Shahan, T. A. Temporal contingency. *Behav. Processes* **101**, 89–96 (2014).
17. Namboodiri, V. M. K. *et al.* Single-cell activity tracking reveals that orbitofrontal neurons acquire and maintain a long-term memory to guide behavioral adaptation. *Nat. Neurosci.* **22**, 1110–1121 (2019).
18. Lin, S.-C., Gervasoni, D. & Nicolelis, M. A. L. Fast modulation of prefrontal cortex activity by basal forebrain noncholinergic neuronal ensembles. *J. Neurophysiol.* **96**, 3209–3219 (2006).
19. Nguyen, D. P. & Lin, S.-C. A frontal cortex event-related potential driven by the basal forebrain. *Elife* **3**, e02148 (2014).
20. Raver, S. M. & Lin, S.-C. Basal forebrain motivational salience signal enhances cortical processing and decision speed. *Front. Behav. Neurosci.* **9**, 277 (2015).
21. Avila, I. & Lin, S.-C. Distinct neuronal populations in the basal forebrain encode motivational salience and movement. *Front. Behav. Neurosci.* **8**, 421 (2014).
22. Mayse, J. D., Nelson, G. M., Avila, I., Gallagher, M. & Lin, S.-C. Basal forebrain neuronal inhibition enables rapid behavioral stopping. *Nat. Neurosci.* **18**, 1501–1508 (2015).
23. Lin, S.-C. & Nicolelis, M. A. L. Neuronal ensemble bursting in the basal forebrain encodes salience irrespective of valence. *Neuron* **59**, 138–149 (2008).
24. Avila, I. & Lin, S.-C. Motivational salience signal in the basal forebrain is coupled with faster and more precise decision speed. *PLoS Biol.* **12**, e1001811 (2014).
25. Kaplan, A., Mizrahi-Kliger, A. D., Israel, Z., Adler, A. & Bergman, H. Dissociable roles of ventral pallidum neurons in the basal ganglia reinforcement learning network. *Nat. Neurosci.* **23**, 556–564

(2020).

26. Ottenheimer, D. J. *et al.* A quantitative reward prediction error signal in the ventral pallidum. *Nat. Neurosci.* **23**, 1267–1276 (2020).
27. Ottenheimer, D. J. *et al.* Reward activity in ventral pallidum tracks satiety-sensitive preference and drives choice behavior. *Sci Adv* **6**, (2020).
28. Schultz, W. Dopamine reward prediction-error signalling: a two-component response. *Nat. Rev. Neurosci.* **17**, 183–195 (2016).
29. Gallistel, C. R., Fairhurst, S. & Balsam, P. The learning curve: implications of a quantitative analysis. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 13124–13131 (2004).
30. Colwill, R. M. & Rescorla, R. A. Postconditioning devaluation of a reinforcer affects instrumental responding. *J. Exp. Psychol. Anim. Behav. Process.* **11**, 120–132 (1985).
31. Dickinson, A., Nicholas, D. J. & Adams, C. D. The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. *The Quarterly Journal of Experimental Psychology Section B* **35**, 35–51 (1983).
32. Lin, S.-C., Brown, R. E., Hussain Shuler, M. G., Petersen, C. C. H. & Kepecs, A. Optogenetic Dissection of the Basal Forebrain Neuromodulatory Control of Cortical Activation, Plasticity, and Cognition. *J. Neurosci.* **35**, 13896–13903 (2015).
33. Paxinos, G. & Watson, C. *The rat brain in stereotaxic coordinates*. (Academic Press, 2007).