HHS Public Access

Author manuscript

*Neuron*. Author manuscript; available in PMC 2023 November 16.

Published in final edited form as:

*Neuron*. 2022 November 16; 110(22): 3789–3804.e9. doi:10.1016/j.neuron.2022.08.022.

# Striatal dopamine explains novelty-induced behavioral dynamics and individual variability in threat prediction

**Korleki Akiti**[1], **Iku Tsutsui-Kimura**[1], **Yudi Xie**[1,2], **Alexander Mathis**[1,3,4], **Jeffrey Markowitz**[5,6], **Rockwell Anyoha**[5], **Sandeep Robert Datta**[5], **Mackenzie Weygandt Mathis**[3,4], **Naoshige Uchida**[1], **Mitsuko Watabe-Uchida**[1,7,*]

[1]Department of Molecular and Cellular Biology, Center for Brain Science, Harvard University, Cambridge, MA 02138, USA

[2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[3]The Rowland Institute at Harvard, Harvard University, Cambridge, MA 02138, USA

[4]Swiss Federal Institute of Technology Lausanne, Geneve, CH 1202, Switzerland

[5]Department of Neurobiology, Harvard Medical School, Boston, MA 02115, USA

[6]Wallace H. Coulter Department of Biomedical Engineering, Emory School of Medicine and Georgia Institute of Technology, Atlanta, GA 30322, USA

[7]Lead Contact

## SUMMARY

Animals both explore and avoid novel objects in the environment, but the neural mechanisms that underlie these behaviors and their dynamics remain uncharacterized. Here, we used multi-point tracking (DeepLabCut) and behavioral segmentation (MoSeq) to characterize the behavior of mice freely interacting with a novel object. Novelty elicits a characteristic sequence of behavior, starting with investigatory approach and culminating in object engagement or avoidance. Dopamine in the tail of striatum (TS) suppresses engagement, and dopamine responses were predictive of individual variability in behavior. Behavioral dynamics and individual variability are explained by a reinforcement learning (RL) model of threat prediction, in which behavior arises from a novelty-induced initial threat prediction (akin to “shaping bonus”), and a threat prediction that is

*Correspondence: mitsuko@mcb.harvard.edu (M.W.-U.).

AUTHOR CONTRIBUTIONS

KA, NU and MW-U initiated the project and designed the experiments. KA and YX set up the arena and wrote analysis code. AM and MWM assisted in DeepLabCut setup. JM, RA, and SRD assisted in MoSeq setup. KA trained the mice and collected the data. KA and IT-K performed surgery and histology. KA and MW-U analyzed the data and wrote the paper. KA, IT-K, YX, AM, MWM, NU, SRD and MW-U edited the paper.

Lead contact: (Mitsuko Watabe-Uchida)



DECLARATION OF INTERESTS

The authors declare no competing interests.

learned through dopamine-mediated threat prediction errors. These results uncover an algorithmic similarity between reward- and threat-related dopamine sub-systems.

## eTOC blurb

Using automated analysis of mouse behavior, Akiti et al. find diverse and dynamic novelty exploration patterns including risk assessment, engagement, and neophobia. These behaviors can be explained by a subset of dopamine neurons that treat physical salience as a default threat estimate, thereby causing progressive avoidance of the novel object.

## INTRODUCTION

In the natural world, animals continuously face the problem of deciding whether to approach, avoid, or ignore a novel stimulus. Maladaptation to novelty has been implicated in anxiety, autism and schizophrenia (Baron-Cohen et al., 2005; Hirshfeld-Becker et al., 2014; Jiujias et al., 2017; Kagan et al., 1984; Orefice et al., 2016). Behavioral responses to novelty have been modeled in different ways across fields. Within the field of reinforcement learning, novelty is often thought of as either a rewarding outcome or a predictor of a potential reward, thereby prompting exploration before the first rewards are received (Kakade and Dayan, 2002; Xu et al., 2021). In this way, novelty can be incorporated into existing reinforcement learning frameworks. Similarly, artificial intelligence models have been created that are “curious” or intrinsically motivated (Colas et al., 2019; Oudeyer et al., 2007, 2016; Stout et al., 2005). Some of these models use information gain, a reduction in the difference between the current event and what was expected over time, to define event novelty (Jaegle et al., 2019; Kaplan and Oudeyer, 2007). Notably, while many computational models of novelty capture the neophilic aspects of novelty behavior, they fail to capture the neophobia and the interplay between approach and avoidance in response to novelty, observed in natural novelty responses.

Dopamine regulates reward-related behaviors, and electrophysiology studies have shown that dopamine signals the discrepancy between actual and predicted reward value (Montague et al., 1996; Schultz et al., 1997). In reinforcement learning, dopamine can be used as an evaluation signal to reinforce a rewarding action. However, recent studies have found that some dopamine neurons are activated by novelty (Horvitz et al., 1997; Lak et al., 2016; Ljungberg et al., 1992; Menegas et al., 2017, 2018; Morrens et al., 2020; Schultz, 1998). To incorporate these novelty responses into the reinforcement learning framework, it has been proposed that dopamine novelty response may correspond to optimism or the potential for reward (Kakade and Dayan, 2002).

Although it has been widely assumed that dopamine neurons broadcast reward prediction error signals to a wide swath of targets, recent studies have shown that dopamine neurons projecting to different targets send distinct information (Kim et al., 2015; Lerner et al., 2015; Menegas et al., 2017; Parker et al., 2016). Importantly, the canonical dopamine system – comprising those neurons that project from the ventral tegmental area (VTA) to the ventral striatum (VS) – does not respond to novel stimuli at the population level (Menegas et al., 2017). Recent studies in monkeys also found that dopamine neurons in substantia nigra

pars compacta (SNc) do not respond to novelty per se (Ogasawara et al., 2022), but rather respond to novelty in the context of information seeking for reward (Bromberg-Martin and Hikosaka, 2009). In contrast, recent studies found that dopamine neurons that project to the tail of the striatum (TS) or the prefrontal cortex play a role in task-independent novelty-related behaviors (Menegas et al., 2018; Morrens et al., 2020).

A recent study found that dopamine in TS displays unique response properties (Kim and Hikosaka, 2013; Menegas et al., 2017). TS-projecting dopamine neurons are strongly activated by high intensity or novel external stimuli in the environment (Menegas et al., 2017, 2018), or by salient visual cues, but not by reward (Kim et al., 2015). Functionally, TS-projecting dopamine neurons facilitate avoidance of a threatening stimulus including a novel object (Menegas et al., 2018).

However, it is not clearly understood how dopamine modulates novelty-driven behaviors, as there are several limitations in previous studies. First, previous studies treated novelty-related behavior as a binary choice of either approach (orient, saccade) or avoidance, and often ignored the behavioral complexity, dynamics and individual variability, which is essential to understand the computations underlying novelty responses. Variability in the novelty-triggered behavioral data had been even interpreted as experimental deficits (Corey, 1978). However, individual variability is an important factor to understand the neural computations (Marder and Goaillard, 2006). Second, many previous studies were conducted in constrained environments that limited behavioral choices (Menegas et al., 2017; Morrens et al., 2020; Ogasawara et al., 2022). It has been reported that animals respond differently to novel objects depending on whether the animal is in a small environment (“forced exposure”) or in a sufficiently large enclosure to be able to choose between exploring or totally avoiding a novel object (“voluntary exploration”) (Corey, 1978; Rebec et al., 1997). Third, the definition of novelty has varied across studies. Recent studies emphasize the computational difference between stimulus novelty and contextual novelty: the former refers to the quality of not being previously experienced or encountered, and the latter refers to the “surprise” when what is experienced does not match with what was expected in time and/or context (i.e. prediction error) (Barto et al., 2013; Kumaran and Maguire, 2007; Ranganath and Rainer, 2003; Xu et al., 2021).

In this study, we used machine learning to characterize individual variability in behavioral novelty responses while mice freely explored a novel object placed in a large arena. We subsequently examined the effects of two types of novelty, the first in which a mouse explored a new stimulus (“stimulus novelty”) and the second in which a mouse explored a familiar stimulus in a new location (“contextual novelty”). These different novelty manipulations induced distinct patterns of behavior, which were differentially affected by ablation of TS-projecting dopamine neurons. The diversity and the dynamics of the observed novelty behaviors were well captured by a simple reinforcement learning model, which incorporates the concepts of initial estimation (“shaping bonus”) and uncertainty. We propose that novelty avoidance is a critical defensive strategy in which a novel stimulus causes default estimation of potential threat when the outcome is unknown. Because death or significant injury prevent learning, the brain may have adapted to estimate the degree of threat posed by a novel object through its physical salience, signaled by dopamine in TS.

# RESULTS

## Novelty triggers diverse behaviors with stereotypical risk assessment response

We designed an open arena novelty exploration paradigm (Figure 1). Mouse movements were captured using an overhead camera that recorded four channels: three color channels (RGB) and one channel for depth (Microsoft Kinect). DeepLabCut (Mathis et al., 2018) was used to track the nose, ears, and tail base of the mouse (see Methods). On the first day of novelty (N1) when the mice were first encountering the object, mice exhibited diverse behaviors; some spent more and some spent less time within the object area compared to habituation days (Figure 1B). The observed diversity was not random noise because time spent near the object (see Methods) in each individual was strongly correlated across sessions (Figure 1B). Novel object approach frequency and approach bout duration also varied across mice, although both of these parameters co-varied within a given mouse (Figure 1C–D).

Close examination of nose and tail trajectories revealed that during the first several approach bouts mice oriented themselves to face the object (Figure 2). As a result, when the mouse reached the closest point to the object, the closest body part was always the nose, not the tail (Figure 2C, N1). These data suggest that the novelty response characterized by “approach with tail behind” is unique to early interactions with a novel object.

To quantify this prominent novelty-related behavior, we classified approach bouts based upon orientation, which revealed that every mouse approached the object with the tail behind in the first 10 min of the first day of novelty (Figure 2D, n=26 animals). The frequency of approach with tail behind decreased over time (Figure 2D). Over the course of the first day, some mice started to expose their tails to the object, while some mice did not expose their tails to the object during entire sessions (Figure 2E).

Thus mice exhibit a robust and stereotyped response at the beginning of interactions with a novel object, one that resembles a form of behavior described as “risk assessment” (Blanchard et al., 1991; Gottlieb and Oudeyer, 2018; Kidd and Hayden, 2015). In contrast, post-assessment behaviors were diverse, with individual animals exhibiting a wide spectrum of approach or avoidance behaviors. We operationally refer to post-assessment approach as “tail exposure engagement,” to distinguish it from risk assessment.

## Post-assessment engagement is suppressed by stimulus novelty

Initial encounters with a novel object inevitably include both stimulus novelty (as the object has not yet been encountered) and contextual novelty (as the object has not been encountered in any context). In order to understand that the stimulus is novel, the brain has to search its stored memory of all objects encountered in the past (Barto et al., 2013). To understand that, in addition, the object is unexpected in the current context, the brain has to compare the current state with the predicted state (Ranganath and Rainer, 2003).

To separate the impact of different kinds of novelty on behavior, we incorporated object pre-exposure into our behavioral paradigm (Figure 3), which dramatically changed the animal’s reaction to the object in test sessions. As shown above (Figures 1–2), some mice

spend more time near a novel object on N1, while others spend less (Figure 3A, left). In contrast, mice consistently approached an unexpected familiar object (Figure 3A, middle). As a population, mice with an unexpected familiar object spent significantly more time near the object than mice with a novel object (Figure 3A, right); they exhibited limited tail behind approach and quickly switched to tail exposure (Figure 3B). Mice interacting with a novel object used tail-behind approach significantly more frequently than mice with an unexpected familiar object and used tail exposure approach significantly less frequently (Figure 3B–C).

Our observation that tail-behind approach was consistently observed at the beginning of N1 in both groups (Figure 3B and D) suggested that risk assessment behavior is driven by unexpectedness, not specifically by stimulus novelty. However, in response to an unexpected familiar object, mice exhibited a quick transition to approach with tail exposure (engagement), suggesting that stimulus novelty suppresses engagement.

## Ablation of TS-projecting dopamine neurons biases post-assessment behavior towards approach

To understand the computational role of dopamine in TS in novelty-driven behaviors, we performed ablation of TS-projecting dopamine neurons with 6-hydroxydopamine (6OHDA) (Figure 4, Figure S1). Consistent with our previous study (Menegas et al., 2018), animals with ablation of TS-projecting dopamine neurons spent more time near a novel object than animals with injection of a vehicle (Figure 4B) and showed longer duration of approach bouts (Figure S2). When analyzing risk assessment and engagement, all ablation mice as well as sham-lesioned animals expressed approach with tail behind in early periods of N1 (Figure 4C, left). After risk assessment, more ablation mice showed transition to tail exposure approach, resulting in higher frequency of tail exposure as a population (Figure 4C–D).

These results demonstrate that ablation of TS-projecting dopamine neurons increased approach with tail exposure, i.e. premature transition to engagement, suggesting that intact dopamine in TS suppresses post-assessment engagement.

## Behavioral segmentation of novelty-driven behaviors

So far, we classified approach types by focusing on animal’s tail position relative to nose. To segment behavioral responses to novel objects into constituent components, we next analyzed the same data using MoSeq (Wiltschko et al., 2015), an unsupervised machine learning-based behavioral characterization method, that identifies behavioral motifs or “syllables” from depth imaging data (Figure 5). We noticed that some syllables were overrepresented near the time of retreat. One syllable stood out (Figure 5B, syllable 79, purple) in both the novel object mice and the sham mice. To examine whether any of the syllables were frequently and specifically expressed in different novelty conditions, we first identified a set of syllables that was both highly used and enriched in the novel or unexpected familiar object condition (see Methods, Figure S3). We found that the identified syllables 79 and 14 were highly enriched at the time of retreat compared to the whole session, nearly always occurring during approach with tail behind rather than with tail exposure (Figure 5C, usage was 46.3% and 22.9% of all approach with tail behind (n=684)

for syllables 79 and 14, respectively). Interestingly, syllable 79 was expressed just before the time of retreat and was reliably followed by syllable 14 (14 follows 79, 71.3%±18.9 of usages, mean±SEM, n=17 sham animals, Figure 5F, left).

Visual inspection of the videos (Video S1, Video S2) and video clips (Figure 5D) revealed that syllable 79 represented a “cautious approach” behavior and that syllable 14 represented a “cautious retreat” behavior. These results indicate that cautious approach and retreat are linked, and together make up risk assessment behavior. Thus, both syllables enriched in the novel object condition were related to risk assessment behavior, which is consistent with our observations made through body part tracking (DeepLabCut) demonstrating that approach with tail behind is more pronounced with a novel object (Figure 3).

Consistent with the temporal dynamics of risk assessment characterized above, syllables 79 and 14 showed a gradual decay in usage (Figure 5E). Interestingly, both syllables were also expressed more frequently in sham mice compared to ablation mice (Figure 5E, Figure S3, sham vs ablation, p=0.010, syllable 79; p=0.030, syllable 14, K-S test). Thus, ablation of TS-projecting dopamine neurons decreased both novelty responses and usage of risk assessment syllables 79 and 14, although our manual classification using DeepLabCut could not detect the small difference (Figure 4D).

Our finding that the expression of both syllables 79 and 14 were decreased in ablation mice indicates that TS dopamine impacts both cautious approach and retreat behaviors. This is surprising because if approach and retreat are opposing behaviors, and dopamine in TS reinforces only retreat, ablation of TS-projecting dopamine neurons should predominantly affect retreat. However, the specific syllables associated with approach and retreat were both affected by ablation. We next compared transition from syllable 79 to 14 in sham and ablation animals. Transition from syllable 79 to syllable 14 was similarly high in both animal groups (Figure 5F), indicating that choice of retreat types, characterized by a combination of syllables 79 and 14, was already determined before approach. Ablation of TS-projecting dopamine neurons decreased risk assessment, characterized by a specific posture of approach-retreat, but did not change the structure of risk assessment behaviors, characterized by the sequence of unique syllables.

## TS dopamine response to novelty reflects individual variability in behavior

To better understand the role that TS dopamine plays in novelty behavior, we monitored dopamine release in TS using fiber fluorometry with a dopamine sensor, GRAB-DA2m (Sun et al., 2020)(Figure 6, Figure S4). Consistent with our previous observations of dopamine axon calcium in TS (Menegas et al., 2018), we observed dopamine release in TS around the time of retreat onset when animals were at the closest point from an object, but not at the start of approach or at the end of retreat (Figure 6A), consistent with the idea of risk assessment or evaluation.

As described above (Figures 1–2), behavioral responses to novelty were variable across animals. Interestingly, the dopamine responses to a novel object were also variable (Figure 6B). Further, mice with high average TS dopamine responses on N1 tended to spend less time near the object (Figure 6C, left), showed less frequent tail exposure (Figure 6C, second

from left), and were slower to transition to the first approach with tail exposure (Figure 6C, right). These correlations held true even if we considered the same number of approach bouts in each analysis (Figure S4C). Thus, the individual variability of dopamine responses corresponded to individual variability in behavior.

On trial-by-trial basis, dopamine responses were significantly correlated with current and next approach types (Figure S4D). Dopamine responses were higher during early risk-assessment phase before the first approach with tail exposure (phase 1) than the late engagement phase after it (phase 2) ($p=0.0059$, n=12 animals, paired t-test, Figure 6E, Figure S4E–F). However, within phase 2, dopamine responses were similar between approach types ($p=0.90$, n=12 animals, paired t-test, Figure 6F). After normalizing for trial number dopamine responses were still correlated with the next approach type, but were no longer correlated with the current approach type (Figure S4E).

Taken together, our recording results reveal that dopamine release in TS correlates with approach types, with smaller responses correlating with individual engagement. However, the specific level of dopamine release in TS was not correlated with the current approach type after normalizing for trial number or within phase 2, suggesting that acute dopamine concentration in TS does not fully explain retreat movement in this paradigm.

## Reinforcement learning model with a shaping bonus and uncertainty for novelty response

We sought to develop a simple model to understand how dopamine signals algorithmically relate to novelty-driven behaviors. In standard reinforcement learning models, dopamine is typically modeled as temporal difference (TD) error. This is the difference between reward predictions (or values) of adjacent states, which can be used as a teaching signal for incremental learning of reward predictions (Sutton and Barto, 2018). Using similar logic, we first modeled simple threat prediction learning with TD error (Figure 7). In this model, trials are denoted by bouts of approach towards and sampling of the object. We added ‘threat’ at the time when an agent reached the object (‘object location’ hereafter; Figure 7, far left), instead of adding reward as in reward prediction learning. Threat prediction is used to determine immediate behavioral choice by comparing prediction with a constant threat threshold. If the current threat prediction is lower than the threat threshold, an agent will engage the novel object. If threat prediction is higher than threat threshold, an agent will avoid the object (Figure 7, far right).

In this model, TD error shows a positive response at the object location, which gradually decreases over many encounters (Figure 7, second panel from right). The decrease of TD error is solely because threat is more predicted, thus generating a smaller prediction error, but the level of threat assigned to the object is kept constant (Figure 7, far left). We then examined how threat predictions developed near the object location. Threat prediction before an agent reaches the object location gradually increased over multiple encounters and eventually plateaus (Figure 7, far right). Because the threat threshold is a constant, the increase of threat prediction translates into a behavioral change from approach to avoidance (Figure 7, far right). While increasing threat prediction explains the later avoidance exhibited by some animals in the novel object group, this explanation is inconsistent with our

observation that some animals eventually showed engagement. Further, it does not explain why animals engage with familiar objects if the object is threatening.

We previously found that TS dopamine responses to a novel stimulus decayed when not associated with an outcome, whereas this decay slowed when it was associated with an outcome, especially a threatening outcome (Menegas et al., 2017). In this case, a novel stimulus can be interpreted as a threat-predicting cue instead of unconditioned threat stimulus. We therefore modeled threat learning with a positive default value of threat prediction assigned to a novel object, similar to a “shaping bonus” (Kakade and Dayan, 2002). A fixed value for the shaping bonus functions as a preliminary, initializing value of threat prediction, which speeds up (‘shapes’) but does not distort eventual learning. In our model, an agent would eventually learn no outcome (no threat) associated with a novel object, but in the meantime, threat prediction and behaviors would be shaped by the initial estimation of threat prediction.

We examined the dynamics of TD errors and threat prediction using different levels of shaping bonus (i.e. initial threat prediction level) (Figure 8). The shaping bonus was applied at the object location to model a tentative guess of threat prediction according to the sampled sensory features without knowing the ultimate outcome. Threat prediction at the object location was defined by the shaping bonus (Figure 8A, fourth column, cyan) and gradually decreased over trials to 0 (Figure 8A fourth column at time 10), because the actual outcome is nothing. In other words, the agent’s initial guess of threat prediction associated with the sensory features was wrong and subsequently updated by learning (Figure 8A, third column).

In the meantime, the threat prediction near the object initially increases because of positive TD errors caused by a shaping bonus, then decreases afterwards and eventually becomes 0 after learning has finished (Figure 8A, fourth and far right columns). Across different conditions as the shaping bonus increases, the peak of the threat prediction near the object increases, whereas the time-course is similar (Figure 8A, far right). The concave shape of threat prediction development near the object explains approach agents who eventually engage with a novel object (Figure 8A, second row), and avoidance agents who first approach but ultimately avoid (Figure 8A, third row, see below for termination of learning with avoidance). Differences in the level of shaping bonus can thus produce different patterns of behavior throughout learning (Figure 8C).

However, animals do not choose behaviors based solely on threat prediction level. Even if their estimate of potential threat is low, they should be still cautious if the estimation is uncertain. We therefore added uncertainty of threat prediction to the model as another determinant of behavior. To implement uncertainty in a principled manner, we used a Kalman filter to incrementally determine estimation uncertainty (see Methods), and plotted this together with threat prediction (Figure 8A, B). In these examples, threat prediction is plotted with a 95% confidence range.

We find that the uncertainty of threat prediction explains dynamics of risk assessment behaviors. With low initial estimation of threat, uncertainty of threat prediction is high

at the beginning, inducing risk assessment behaviors, but the uncertainty quickly decays and allows a fast switch to engagement (Figure 8A, first row). Similarly, an unexpected familiar object causes an initial risk assessment because of threat uncertainty, but does not induce avoidance because the initial estimation of threat prediction with the object features is already canceled out by learning during pre-exposure (Figure 8A, bottom row). On the other hand, with high initial estimation of threat, uncertainty is high at the beginning, and then threat prediction increases, causing longer risk assessment (Figure 8A, second row). If threat prediction gets bigger than a threshold, agent chooses to avoid. Once it avoids, it loses a chance to further learn threat prediction that would eventually become 0, which results in persistent avoidance (neophobia) (Figure 8A, third row). Thus, the degree of shaping bonus may determine whether an agent becomes neophobic or not.

The shaping bonus in this model is determined by the initial responses of dopamine in TS to an object, and initial responses vary by individual. Our previous studies found that responses of TS-projecting dopamine neurons are monotonically modulated with the physical salience (intensity) of an external stimulus in the environment (Menegas et al., 2018). Thus, representation of physical salience in TS dopamine will determine the shaping bonus in this model, which in turn facilitates development of threat prediction and affects future actions. Taken together, these results suggest that behavioral engagement with a novel object is well captured by a reinforcement learning model with a shaping bonus, one in which threat prediction builds up according to representation of physical salience of the object in TS dopamine. By changing the level of shaping bonus, which can be inferred from the level of TS dopamine, the model predicts the diverse and dynamical patterns of behaviors observed across individuals and experimental conditions.

As an alternative model, we next modeled that TS actively promotes assessment by signaling prediction of prediction errors (“salience”), while too much of salience causes avoidance (Figure S5, see Methods). A simple TD learning was applied. We found that salience near an object initially increases and then decreases as an agent learns an object. By setting a threshold for avoidance, this model also predicts diverse and dynamic behaviors depending on TS dopamine.

## DISCUSSION

In this study, we propose a reinforcement learning model that captures behavioral dynamics and variability in response to novelty. We were led to this model by examining novelty-induced behaviors in freely-moving animals using supervised (DeepLabCut) and unsupervised (MoSeq) machine learning tools. These approaches demonstrate that all mice initially exhibit risk assessment behaviors toward a novel object, followed by engagement or avoidance. Behavioral syllables that are enriched at the beginning of a novel object exploration correspond to cautious approach and cautious retreat, which together constitute a set of risk assessment behavior. Thus, our application of machine-learning-based analysis methods allowed us to identify distinct behavioral motifs that are dynamically driven during an encounter with a novel object.

The observed distinct approach behaviors depart from the previous studies that categorized novelty-induced behaviors merely by two opposing choices (approach versus avoidance) along a single dimension. By distinguishing the approach types, we found that stimulus novelty and dopamine in TS specifically suppress post-assessment engagement, but not risk assessment. We constructed a simple temporal difference (TD) learning model by incorporating an initializing value (shaping bonus) and uncertainty of threat prediction. In this model, TS dopamine, which conveys a threat prediction error, gradually builds up threat prediction over multiple encounters with a novel object. This in turn suppresses the transition from the risk assessment phase to post-assessment engagement, causing neophobia in extreme cases. Thus, in contrast to classical animal behavior models of novelty, neophobia can be caused by development of threat prediction rather than novelty detection per se. As the object turns out not to be threatening, threat prediction gradually decreases which models habituation. In this way, the model captured not only the temporal dynamics of novelty responses, but also individual variability in the behaviors. Importantly, we found that variability in TS dopamine responses corresponded to individual variability in behavioral responses, providing a neural readout of shaping bonus for threat learning. Together, our findings provide insights into the computations and neural mechanisms that may underlie the dynamics of novelty-induced behaviors, including neophobia.

## Shaping bonus and neophobia

Novelty drives both immediate behavioral responses and learning. Various computational models incorporate novelty components to understand optimal strategies and animal behaviors, because the generation of appropriate novelty responses has been linked to behavioral strategy and learning in daily life (Jaegle et al., 2019; Kakade and Dayan, 2002). While most computational models have focused on the approach aspect of novelty responses, our study has extended these ideas to model approach suppression by incorporating a shaping bonus and uncertainty into a reinforcement learning framework.

Learning an appropriate action is often difficult, because the action is too complicated to learn at once and because an action and its outcome are too temporally separated to easily establish causality. Therefore, in operant conditioning, it is often the case that behaviors are “shaped by making the contingencies of reinforcement increasingly more complex” (Skinner, 1975). In machine learning, some powerful learning models are often slow. To make learning more efficient and fast, some models have copied the idea of shaping from psychology by adding an extra reward (“shaping” or sometimes called an intrinsic reward) at an intermediate step for learning of longer sequential choices (Ng et al., 1999; Singh et al., 2010). However, adding an extra intermediate reward distorts the eventual learning; an agent might learn to acquire only the mid-point reward, which prevents from learning from the actual reward in the future. To overcome the problem of learning distortion, a specific form of shaping (“potential-based shaping”) has been proposed (Ng et al., 1999). In this method, instead of adding an extra reward, reward expectation is added at an intermediate step to preserve original reward function, but still “shapes” an agent’s actions and learning (Wiewiora, 2003). As a consequence, reward prediction of a state is initialized with a positive value even before an agent has visited that state.

Optimal initialization of control systems plays a critical role not only in machine learning but also in animal behaviors. For example, animals can avoid some threatening stimuli using species-specific defensive systems even if they have never encountered them. These phenomena can be interpreted as an initialization of threat prediction with a pre-programmed value. In addition to these pre-programmed mechanisms, the initializing value could in principle be set by experiences in other states without visiting the actual state. Such flexible initialization is critical for efficient machine learning ("smart initialization" (Simsek et al., 2011)), and for behavioral choices in daily life, where agents/animals continuously face novel states. Rather than starting from uniform estimation over all states, an initial guess (generated via evolution and/or generalization) can help to quickly learn more accurate estimation.

In reinforcement learning, approach to novel objects or cues is often modeled using a "novelty bonus" or "shaping bonus". We adapted this approach to model avoidance of a novel object. Our model differs from previous animal behavior models of novelty where fear is simply a decaying function with novelty (Blanchard et al., 1991; Gordon et al., 2014; Halliday, 1966; Hogan, 1965; Hughes, 1997; Lester, 1967; Montgomery, 1955; Thorpe, 1956) in that it predicts that threat prediction first builds up and then (potentially) decays. These dynamics explain a variety of observed behavioral patterns. We also incorporated uncertainty of the threat prediction into our model, thereby accommodating threat predictions ranging from risk assessment to engagement. Interestingly, we found a unique phenomenon specific to threat learning. Once an agent learns that the object is threatening, an agent avoids the object entirely and loses a chance to further learn. As a consequence, the agent gets trapped in an avoidance state. Thus, our model changes the way we interpret neophobia. Neophobia may not be simply driven by abnormal novelty detection per se, but instead forms dynamically in two steps. Uncertainty of safety induces initial risk-assessment, which is followed by a learning process about which objects should be avoided.

Since neophobia was thought to be linked to novelty, brain areas engaged during neophobia have been proposed to be involved in novelty detection. In this study, we found that TS dopamine plays a role in neophobia. While we cannot exclude possibility that TS dopamine is involved in novelty detection, TS dopamine likely signals the physical salience (such as intensity) of external stimuli. Activity of dopamine in TS is initially correlated with the intensity of novel stimuli (Menegas et al., 2018) and then gradually decays depending on associated future events (Menegas et al., 2017). Thus dopamine responses in TS, instead of detecting novelty, are initialized depending on stimulus salience, and then responses are adjusted afterwards. Our model further predicts that TS dopamine excitation with positive initialization ('potential threat' associated with strong physical salience) is used as an evaluation signal for learning of threat prediction at an earlier time point (before approach), which in turn prevents animals from approaching a potential threat. In this way, TS dopamine system uses physical salience of a stimulus as a default value of threat prediction to shape defensive behaviors even before animals learn the exact threat level. Hence, neophobia may be caused by abnormal threat prediction due to general sensitivity to sensory stimuli, rather than aberrant novelty detection.

Why, then, do animals avoid a novel salient stimulus in the first place? A recent series of studies found that in appetitive situations, the taste of food is not an ultimate outcome but instead functions as a prediction of nutrients, which are the ultimate consequence of eating (Fernandes et al., 2020; Han et al., 2018; Tellez et al., 2016). From these results, Dayan proposed that taste is a kind of shaping, an initial guess for value of eating, which can be updated according to an actual outcome, i.e. nutrients (Dayan, 2021). In this framework, dopamine responses to food rewards (taste, or odor (Morrens et al., 2020) ) are tentative feedback based on shaping bonus, but not ultimate reward outcome, to facilitate learning. We can interpret our threat prediction data by analogy to the idea in appetitive value (Figure S6). Similar to well-known pre-programed threats such as looming stimuli and predator odors, physical salience of stimuli may help animals to estimate threat without actual experiences. While many salient stimuli end up being non-threatening, caution against exploring high intensity novel stimuli may be lifesaving. Physical salience can be easily and quickly computed and easily generalized. Therefore, animals may routinely use physical salience as an initial guess of a potential threat for an immediate action and learning, because learning threat only from ultimate outcomes such as pain, injury and death may come at a high cost. Thus, the idea of shaping can be broadly applicable, and dopamine neurons with distinct activities can share a common framework.

## Diversity of dopamine neurons

While the role of dopamine in reward prediction has been relatively established (Eshel et al., 2013; Glimcher, 2011; Schultz, 2016; Watabe-Uchida and Uchida, 2018), our knowledge of functional diversity of dopamine neurons is still incomplete (Cox and Witten, 2019; Watabe-Uchida and Uchida, 2018). In particular, it is not yet clear whether non-canonical dopamine signals can be understood in the similar theoretical framework or algorithm as those in reinforcement learning theories. In our previous studies, we found that TS-projecting dopamine neurons do not signal rewards but respond to a set of external stimuli in the environment, especially high intensity or novel stimuli (Menegas et al., 2017, 2018), and play a role in avoidance of them (Menegas et al., 2018).

Based on precise observation of behaviors and dopamine signals in response to novelty, we have obtained a clearer view on how TS dopamine functions during novelty exploration. First, it should be noted that, unlike previous experiments (Cohen et al., 2012; Menegas et al., 2017; Schultz et al., 1997; Tsutsui-Kimura et al., 2020), our work involves animals freely interacting with an environment. Nonetheless, discrete approach-retreat bouts in our novelty paradigms can be regarded as being equivalent to “trials” in more structured behavioral paradigms, albeit with a critical difference in that the animal can control “task” structure. Our results support the possibility that non-canonical dopamine signals found in TS work as an evaluation signal even in a naturalistic setting, in a manner similar to canonical dopamine signals observed in many structured tasks (Cohen et al., 2012; Glimcher, 2011; Schultz, 2015) or during social interactions (Dai et al., 2021; Gunaydin et al., 2014). Further, dopamine in TS, while signaling totally different information from canonical dopamine, may facilitate salience prediction (threat prediction if salience is too strong) in a similar manner that canonical dopamine facilitates reward prediction.

Together, our results suggest a possibility that even if information contents are diverse, the function of dopamine neurons can be understood within the common framework of reinforcement learning including an idea of bonuses for fine tuning.

## STAR Methods

### RESOURCE AVAILABILITY

**Lead contact—**Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Mitsuko Watabe-Uchida (mitsuko@mcb.harvard.edu).

**Materials availability—**This study did not generate new unique reagents.

**Data and code availability—**Matlab code files are available on GitHub (https://github.com/ckakiti/Novelty_paper_2021).

Video tracking and dopamine fluorometry data are deposited at Dryad (doi:10.5061/dryad.41ns1rnh2).

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

**Animals—**78 adult male and female mice were used. Behavioral experiments were performed on C57BL/6J mice (Jackson Laboratories, RRID: IMSR_JAX:000664), aged 9–17 weeks, on the dark cycle of a 12-hr dark/12-hr light cycle (dark from 7:00 to 19:00). Behavioral tests and recordings were conducted between 8:00 and 18:00. Animals were group-housed until testing or surgery, then individually housed throughout testing. All procedures were performed in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and approved by the Harvard Animal Care and Use Committee.

### METHOD DETAILS

**Behavioral apparatus—**To assess naturalistic behaviors in mice, an open-field arena was developed that allowed the recording of free movement (see Key Resources Table for parts list). Mice were able to explore freely in a 60cm by 60cm flat arena, either empty or containing a single novel object in one corner. To record movement, a single camera was mounted on a beam ~70cm above the floor of the arena. A bright white LED light (Westek Indoor Outdoor White LED Rope Light) illuminated arena from above.

**Experiment workflow—**Before start of experiment, mice were separated and individually housed at least 1 day in advance. Once separated, mice were then handled for 30 minutes per day for 3 days (see Handling). For a novel object/an unexpected familiar object tests, mice were then pre-exposed to the test object (or dummy object) in their home cage for 30 minutes per day for 4–7 days. Test objects were either legos (Mega Bloks First Builders 80-piece Classic Building Bag, 72 mice) or rubber dog toys (Kong Classic dog toy size M, 6 mice). Brand new test objects were used at the start of each set of mice (fresh out of packaging). Dummy objects were plastic coconut cups (Shindigz 16-oz Coconut Cups,

5.5-in tall). For each animal, the same object was used for the duration of the experiment (1 object per animal, each animal's object was wiped with ethanol after every day). A novel object group and a sham surgery group were pooled for Figure 1.

**Handling:** Handling consisted of weighing mice (on first day) and scooping mice into a transport box. This scooping was to acclimate a mouse to the way they would later be transferred from the behavioral arena back to their home cage. To scoop a mouse, the experimenter would hold a takeout box in a corner of the home cage, laying sideways with opening facing center of cage. The experimenter would wait until the mouse approached and walked into the box before lifting the box up and tilting it gently upright. Then the box would be tilted back sideways, replaced onto the floor of the cage, and the mouse would be allowed to return to cage. If the mouse did not voluntarily approach the box within 10 minutes, the opening of the box would be moved closer to the mouse to encourage entry. Sessions lasted for 30 minutes or until the mouse was scooped at least 5 times, whichever occurred first.

**Pre-exposure:** During pre-exposure, one object was placed in each mouse's cage according to experimental condition (test object or dummy object). Each mouse's object was kept consistent across pre-exposure days, with each object being wiped with ethanol between days. Sessions lasted 30 minutes per day for 7 days. During session, mouse would be allowed to explore object freely within home cage (including touching, moving, etc). Pre-exposure sessions occurred in dimly lit rooms.

**Habituation:** During habituation, animals were placed in empty behavioral arena and allowed to explore freely. Mice were transferred from their home cage into the behavioral arena and transferred out of the arena by scooping with a takeout box. Behavior was recorded with a single overhead camera (Xbox Kinect; see Key Resources Table for materials list). Habituation sessions lasted 25 minutes per animal per day for 2 consecutive days. Mice were run in the same order each day (order determined randomly at the beginning of the experiment, and held constant for the rest of the experiment). If arena was soiled at the end of a session, feces would be removed and floor of arena would be spot cleaned using ethanol-soaked wipes before the next session began. Between rounds of experiments, arena was thoroughly cleaned and base of arena was wiped down with odorless eliminator (Ah! Products All Clear Odorless Odor Eliminator).

**Novelty testing:** Novelty testing sessions consisted of animals exploring a single novel object within the behavioral arena. Object was placed in the corner of a behavioral arena (taped to floor to prevent animal from moving it, ~12–15cm from either wall). Sessions lasted for 25 minutes per animal per day for 4–12 days, and mice were run in the same order as habituation each day. One object was used per animal for duration of experiment and the objects were not shared between animals. Before each session, object would be submerged in soiled bedding (mixture of bedding from each mouse's cage in current round, 6 animals) and wiped off with dry kimwipe to remove excess bedding dust. Objects were wiped with ethanol after each day and allowed to air out overnight before use.

## Video recording and analysis

**Recording:** An Xbox One Kinect camera (Key Resources Table) was mounted 70cm above the behavioral arena. Mice were videotaped with four channels: three color channels (RGB, 15fps) and one depth channel (30fps). The RGB video was used to locate body part locations (DeepLabCut) and the depth video was used to segment behavior (MoSeq). Data was saved using custom recording software (Wiltschko et al., 2015). Analysis code and instructions for running them are deposited on GitHub (https://github.com/ckakiti/Novelty_paper_2021).

**DeepLabCut analysis:** For body part tracking, we used DeepLabCut version 1.0 (Mathis et al., 2018). Separate networks were used for different experimental settings: namely for mice without fiber implants (network A) and mice with fiber implants (network B). Both networks were run using a ResNet-50-based neural network (He et al., 2016; Insafutdinov et al., 2016) with default parameters for 1,030,000 training iterations. We provided manually labeled locations of four mouse body parts within video frames for training: nose, left ear base, right ear base, and tail base. For network A: We labeled 1760 frames taken from 64 videos. For network B: We labeled 540 frames taken from 17 videos. For both networks, 95% of labeled frames were then used for training.

After running DeepLabCut on each video file, we processed the output files (csv array with x/y coordinates and likelihood values for each body part). First, we trimmed the early frames that had low (<10%) likelihood values, indicating that the mouse was not present in the arena yet, or they had poor tracking. We then corrected “jumps” in tracking, defined as a >15cm/frame change in Euclidian distance. Points identified as jumps were replaced by the mean of the previous frame and the following frame. Jumps were corrected separately for each body part (nose, left ear, right ear, and tail). Trajectories for each body part were then smoothed using a lowest moving average filter (5 points, default).

A body part was determined to be “near” the object if it fell within a radius of 7cm (Euclidean distance) from the center of the object. An approach bout is defined as either the nose or tail entering near the object, and the end of this bout is determined when the nose and tail are no longer near the object. Habituation sessions did not have an object present; therefore the area of analysis was chosen based on the position where the object would be in later sessions. This radius was chosen to not be too large and include edge walking (since the object was placed near the corner) but also not to be too small and fail to capture enough of the animals’ trajectory. These approach bouts can be further broken down into whether the nose was closer to the object than the tail for the entire bout (approach with tail behind) or whether the tail was closer at some point (approach with tail exposure). Frequency of tail behind or tail exposure were calculated based on the number of bouts with tail behind approach versus tail exposed approach. Retreat timing was determined to be the closest point of the nose relative to the object before the mouse moves away. Previous studies have used “stretched-attend” posture to detect risk assessment (Blanchard et al., 1991; Fanselow, 1994).

**MoSeq analysis:** Raw imaging data was collected from the depth camera, pre-processed (filtered, background subtracted, and parallax corrected), and submitted to a machine learning algorithm that evaluates the pose dynamics over time (Wiltschko et al., 2015). During video extraction (moseq2-extract), 900 frames were trimmed from the beginning of the video to correct for time between when video was started and when the mouse was placed in arena. During model learning (moseq2-model), a hyperparameter was set to the total number of frames in the training set (kappa=2,711,134, 52 sessions, 52 animals). This exceeds the recommended >=1 million frames (at 30 frames per second) needed to ensure quality MoSeq modeling.

To align syllables to retreat timing, MoSeq data was aligned to DeepLabCut timeframes. This alignment was necessary because the depth and rgb videos have different frame rates (depth=30fps, rgb=15fps; timestamps are saved alongside raw data). We first extracted the timestamps and syllables associated with each frame in the depth video (scalars_to_dataframe function; see GitHub repository “moseq2-app”, code available on request: datta.hms.harvard.edu/research/behavioral-analysis/). We then aligned the depth video timestamps to the corresponding rgb video timestamps (custom MATLAB script, see GitHub repository “Novelty_paper_2021”). This alignment was then used to determine which syllables were expressed at each frame in the RGB videos. We then identified retreat timing and the corresponding MoSeq syllable in each RGB video.

MoSeq was first used to categorize postures into a total of 100 syllables using a combined data on novelty day 1 across the 4 experimental groups: novel object, familiar object, control, and ablation. In order to find a set of syllables that was both highly used and enriched in the novel or unexpected familiar object condition, we chose 10 most frequently occurring syllables around the object (−1s to 1s from retreat time) in each of 4 experimental groups, a total of 21 syllables, and compared the frequency of syllables in each animal. We identified 2 syllable (syllables 79 and 14) that were enriched in the novel object group ($p=0.00049$ both for syllables 14 and 79, K-S test). Bonferroni correction was applied to correct for multiple comparison. No syllables were significantly enriched in the familiar object group with this analysis, although we observed multiple syllables that showed the tendency.

**Surgical procedures—**All surgeries were performed under aseptic conditions with animals anesthetized with isofluorane (1–2% at 0.5–1.0 l/min). Analgesia was administered pre- (buprenorphine, 0.1mg/kg, I.P.) and post-operatively (ketoprofen, 5 mg/kg, I.P.). At the time of surgery, mice were 2–4 months old. We used the following coordinates to target injections and implants for tail of striatum (TS): Bregma: −1.5 mm, Lateral: +3.0 mm, Depth: −2.4 mm (relative to dura) (Paxinos and Franklin, 2019).

**6OHDA surgical procedure:** To bilaterally ablate dopamine neurons projecting to TS, we followed an existing protocol (Menegas et al., 2018; Thiele et al., 2012). The following solution was injected (I.P.) to animals at 10 mg/kg:

- 28.5 mg desipramine (Sigma-Aldrich, D3900–1G)
- 6.2 mg pargyline (Sigma-Aldrich, P8013–500MG)

- 10 mL water
- NaOH to pH 7.4

Most animals (weighing ~25g) received ~250 μL of this solution. This was given to prevent dopamine uptake in noradrenaline neurons and to increase the selectivity of uptake by dopamine neurons. After injection, mice were anesthetized as described above. We then prepared a solution of 10 mg/mL 6-hydroxydopamine (6OHDA; Sigma-Aldrich, H116–5MG) and 0.2% ascorbic acid in saline (0.9% NaCL; Sigma-Aldrich, PHR1008–2G). The ascorbic acid in this solution helps prevent 6OHDA from breaking down. Control animals were injected with vehicle ascorbic acid solution. To further prevent 6OHDA from breaking down, we kept the solution on ice, wrapped in aluminum foil, and it was used within three hours of preparation. If the solution turned brown in this time (indicating that 6OHDA has broken down), it was discarded and fresh solution was made. 6OHDA (or vehicle, ascorbic acid solution) was injected bilaterally into TS (200nL per side). Each injection was spread out over several minutes (70–100 nl per minute) to minimize damage to the tissue. Surgeries occurred 1 week before handling.

**Dopamine sensor surgical procedure:** For TS neurons to express dopamine sensor for fluorometry, we unilaterally injected mixed virus solution (AAV for dopamine sensor and tdTomato, 1:1 mixture, 350 nl total) into TS in WT mice. Virus injection lasted around 5 minutes (injection of 70–100 nl per minute), after which the pipette was slowly removed to prevent damage to the tissue. We also implanted optic fibers (400 μm diameter, Doric Lenses, Canada) unilaterally into the TS (one fiber per mouse). Once fibers were lowered, we attached them to the skull with UV-curing epoxy (Thorlabs, NOA81), then waited for 15 min for this to dry. We then added a layer of black Ortho-Jet dental adhesive (Ortho-Jet, Lang Dental, IL). We used magnetic fiber cannulas (Doric Lenses, MFC_400/430) to allow for recording in freely moving animals. We waited for 15 min for the dental adhesive to dry, and then the surgery was complete.

**Histology and immunohistochemistry—**Histology was conducted in the same manner as previously reported (Tsutsui-Kimura et al., 2020). Mice were perfused using 4% paraformaldehyde, then brains were sliced into 100μm thick coronal sections using a vibratome (Leica) and stored in PBS. These slices were then stained with rabbit anti-tyrosine hydroxylase (TH; AB152, EMD Millipore, RRID: AB_390204) at 4°C for 2d to reveal dopamine axons in the striatum, dopamine cell bodies in the midbrain, and other neurons expressing TH throughout the brain. Slices were then stained with fluorescent secondary antibodies (Alexa Fluor 594 goat anti-rabbit secondary antibody, A-11012, Invitrogen, RRID: AB_2534079) at 4°C for 1d. Slices were then mounted in anti-fade solution (VECTASHIELD anti-fade mounting medium, H-1200, Vector Laboratories, CA) and imaged using Zeiss Axio Scan Z1 slide scanner fluorescence microscope (Zeiss, Germany).

### Fluorometry (photometry) recording

**Overview:** Fiber fluorometry signal was recorded from the striatum in mice performing open field novelty behavior tasks (15 animals). Mice were injected either with AAV to express dopamine sensor. After undergoing surgery (details in Surgical Procedures), animals

were allowed to recover for 2 weeks before the start of behavior testing. In the last 3 days of this period, animals were handled (details in Handling). Then animals went through habituation and novelty testing in the arena (described in a previous section). During photometry recordings, a long flexible optic fiber (see Recording section) was attached to connector on the animal's skull which did not impede animal movement.

**Handling:** In addition to weighing and scooping mice in the takeout box, photometry mice also had a patch cord attached and removed once during the session (not connected to laser, no light transmitted). Animal was allowed to briefly move about cage with patch cord attached (~10s) before being picked back up and disconnected from patch cord. Attachment and removal were conducted in same manner that they would be later in behavioral sessions.

**Recording:** Fluorometry recording was performed as previously reported (Menegas et al., 2018; Tsutsui-Kimura et al., 2020). The following describes this established setup: We use an optic fiber to stably access deep brain regions and interfaces with a flexible patch cord (3 m, Doric Lenses, Canada) on the skull. The patch cord simultaneously delivers excitation light (473 nm, Laserglow Technologies, Canada; 561 nm, Opto Engine LLC, UT) and collect dopamine sensor and tdTomato fluorescence emissions. Activity-dependent fluorescence emitted by cells in the vicinity of the implanted fiber's tip (NA=0.48) was spectrally separated from the excitation light using a dichroic, passed through a single band filter, and focused on a photodetector connected to a current preamplifier (SR570, Stanford Research Systems, CA).

During photometry recording, optic fibers on the animal's skull were connected to a magnetic patch cable (Doric Lenses, MFP_400/430) which both delivered excitation light (473 and 561 nm) and collected emitted light. The emitted light was then filtered using a 493/574 nm beam-splitter (Semrock, NY), followed by a 500 ± 20 nm (Chroma, VT) and 661 ± 20 nm (Semrock, NY) bandpass filters and collected by a photodetector (FDS10 X 10 silicone photodiode, Thorlabs, NJ) which is connected to a current preamplifier (SR570, Stanford Research Systems, CA). This preamplifier outputs a voltage signal which was collected by a NIDAQ board (National Instruments, TX) and custom Labview software (National Instruments, TX, RRID:SCR_014325).

Lasers were turned at least 30 minutes prior to recording to allow them to stabilize. Before each recording session, laser power and amplifier settings were individually adjusted for each mouse. First, the laser power was set low enough to avoid bleaching and high enough to detect signal. Then, the amplifiers were set such that the baseline signals recorded through LabView were similar across mice and days (3–6 a.u. at start of session). Behavior and photometry signal were measured simultaneously using Labview software (see Synchronization section below). After each recording session, collected light intensity was measured from the patch cord using a photometer. Light intensity fell within a range of 15–180μW across animals and days.

**Signal analysis:** DA sensor (green) and tdTomato (red) signals were collected as voltage measurements from current pre-amplifiers (SR570, Stanford Research Systems, CA). Green and red signals were cleaned by removing 60 Hz noise with bandstop FIR filter 58–62

Hz and smoothing with a moving average of signals in 50 ms. The global change within a session was normalized using a moving median of 100 s. Then, the correlation between green and red signals was examined by linear regression. If the correlation was significant ($p<0.05$), the fitted red signals were subtracted from green signals. Z-scores were calculated using an entire recording session. Retreat start was defined as the time point when the animal's nose was closest to the object within an approach bout. Only one retreat start was detected in each approach bout to avoid using multiple time points close each other. Approach start was defined as the time point when the distance between the animal's nose and the object started decreasing before each retreat start. Retreat end was defined as the time point when the distance between the animal's nose and the object started decreasing after each retreat start. Responses aligned at a behavioral event were calculated by subtracting the average baseline activity (−3s to −1s before the event) from the average activity of the target window (0–1s after the event). To show overall activity patterns (Figure 6A), the average activity (−3s to −1s before approach start) was used as baseline.

**Synchronization:** In order to match photometry signal to behavior, it was important to synchronize the rgb video and photometry data. To achieve this, an LED was mounted within view of rgb camera such that it appeared in video, but did not overlap the floor of the arena or obscure the mouse. Custom LabView software was programmed to send a short TTL signal for a brief LED pulse every 10s for the duration of recording. TTL pulses and photometry signal were recorded simultaneously. After recording, the timing of LED flashing in the rgb video was determined and matched with the corresponding TTL pulses that had been saved alongside photometry signal. The result is two arrays of the same length: one containing the RGB frame number for each LED flash and the other containing the photometry timestamp for each TTL pulse (i.e. every 10s). The time for other frames were determined by evenly spacing those frames within 10s intervals.

## Modeling

**Reinforcement learning of threat prediction:** We applied the standard formulation of temporal difference (TD) learning (Schultz et al., 1997; Sutton and Barto, 1990) to threat prediction. In standard TD learning models (Sutton and Barto, 2018), an agent predicts the cumulative future rewards, or value. In our TD model, an agent predicts the cumulative future threats (threatening outcomes) to guide its behavior. We note that TD learning algorithm was originally developed for explaining the strength of association in a type of aversive conditioning (nictitating membrane response)(Sutton and Barto, 1987, 1990). There have also been some efforts to generalize TD learning algorithms to predictions of other quantity or outcomes (or "cumulants") than value (Dayan, 1993; Schlegel et al., 2021). Our application of TD learning to threat prediction takes a similar approach to these precedents.

The threat prediction at time $t$ is denoted as $TP(t)$, and is defined by,

$$TP(t) = E\left[\sum_{k=0}^{N-t} \gamma^k \cdot threat(t+k)\right]$$

where $E[\ldots]$ denotes expectation, $threat(t)$ denotes a threatening outcome occurring at time $t$, and $\gamma \in (0,1)$ is a discount factor. The model contained N (N=350) discrete states or timesteps, which constitute an entire bout of novel object exploration, with a novel object occurring upon entering to the 100$^{th}$ state ($t = 10$) (for convenience, we express time $t$ as the number of timesteps divided by 10). For simplicity, we applied a form of state representation called a complete serial compound, in which an agent deterministically traverses each of the 35 states in sequence (Schultz et al., 1997; Sutton and Barto, 1990), without considering avoidance action that would terminate state transitions and, thus, learning (see below).

In the first model (Figure 7), we assumed that a threatening outcome occurred when the animal encountered a novel object (i.e. $t = 10$). That is, the novel object itself is a threat. Thus,

$$threat(t = 10) = c, \quad threat(t \neq 10) = 0$$

where $c$ is a constant (in the Figure, $c = 2$ was used). Threat prediction, TP, was initialized to 0 for all the states before trial 1.

$$TP(t) = 0 \quad for\ all\ t$$

In each trial, the eligibility trace, $e_t$, was initialized to 0 at the beginning of a trial. At each time $t$, TD errors, $\delta$, were computed similar to a standard definition of TD error (Sutton and Barto, 1987) as the difference between the threat prediction at consecutive time steps plus received threats at each time step.

$$\delta = threat(t) + \gamma \times TP(t+1) - TP(t)$$

Eligibility trace, $e_t$, for each state was updated by decaying $e_t$ by the discount factor ($\gamma$) and the eligibility trace parameter ($\lambda$). For the current state, 1 was added.

$$e_t = \gamma \times \lambda \times e_t \quad if\ t \neq current\ state$$

$$e_t = \gamma \times \lambda \times e_t + 1 \quad if\ t = current\ state$$

Threat prediction was updated according to the obtained $\delta$ and $e_t$,

$$TP(t) = TP(t) + \alpha \times \delta \times e_t$$

where $\alpha \in (0,1)$ is a learning rate. Then, an agent moves to the next time step, starting the next iteration of threat prediction. In this model, TD error at object ($t = 10$) is expressed as:

$$\delta(t = 10) = c - TP(t)$$

which is simply threat minus learned threat prediction.

The second model (Figure 8) does not experience an actual threatening outcome but an initializing value (0 to 2) of threat prediction (i.e., shaping bonus Φ) was added to the state containing a novel object ($t = 10$) that gradually decays, to simulate lingering threat prediction until the animal finds out that there is no threat outcome. Thus, before starting the trial 1,

$$\Phi(10 \leq t \leq 34) = c \times decay^{t-10}\ (shaping\ bonus)$$

$$threat(t) = 0$$

We used constant $c$ from 0 to 2, and decay=0.98 in the Figure 8 Different levels of $c$ yielded different time-course of threat prediction and prediction error in this model. Since Φ is an initializing value of threat prediction, threat prediction can be expressed as:

$$TP = \Phi + \mathrm{TPl}$$

where TPl denotes learned component of threat prediction. Iteration of threat prediction was performed similarly to the model 1.

$$TP(t) = TP(t) + \alpha \times \delta \times e_t$$

Since shaping bonus is fixed across trials, the learning rule can be also expressed as:

$$TPl(t) = TPl(t) + \alpha \times \delta \times e_t$$

In this model, TD error is expressed as:

$$\delta = \gamma \times TP(t+1) - TP(t)$$

because there is no actual threat in any time step.

In all simulations, the learning rate α, the discounting rate γ and the parameter for eligibility trace λ were fixed to 0.02, 0.98, and 0.9, respectively, without model exploration.

For broader application, threat prediction at the decision point can be interpreted as prediction associated with an “object”, whereas the shaping bonus is linked to physical salience of sensory features. While the shaping bonus was applied at the object location (thus representing proximal sensory features including visual details, odors and textures) to simplify the model, shaping bonus can be applied to multiple time points to accommodate other sensory features at a distance. Of note, different from shaping of food approach, which also shapes learning itself by promoting visits, shaping of threat avoidance, which is associated with avoidance of an object, does not promote threat learning itself.

**Uncertainty:** Uncertainty of threat prediction (estimation uncertainty), $pp(n)$, in each trial $n$ was determined incrementally using the following equation (Kalman filter):

$$K = \frac{pp(n)}{pp(n) + pm(n)}$$

$$pp(n+1) = (1 - K) \times pp(n)$$

where $pm$ is a measurement uncertainty. The model used standard normal distribution for estimation (threat prediction) in trial 1, and measurement (actual threat) in all trials, so that both variance $pp(1)$ and $pm\,(n)$ was set to 1.

While we used a frequency-based simple Kalman filter to compute uncertainty, other methods – such as those based on probability distributions over threat levels – could be used to compute uncertainty. While a recent study analyzing single neuron activity found evidence supporting distributional reinforcement learning in the canonical dopamine neuron population (Dabney et al., 2020; Lowet et al., 2020), whether the distributional code observed in dopamine activity is actually used in biology, and whether similar diversity consistent with distributional reinforcement learning is observed in TS-projecting dopamine neurons remain to be clarified.

**Behavioral choice:** Behavior (risk assessment, engagement and avoidance) was chosen every time the agent entered the state near the object ($t = 8$), according to the threat prediction near the object, $TP(t = 8)$ and uncertainty, $pp(n)$, compared to a threat threshold, $thresh$.

$$risk\ assessment\ if\ TP(t=8) - sqrt(2 \times pp(n)) < thresh < TP(t=8) + sqrt(2 \times pp(n))$$

$$engagement\ if\ TP(t=8) - sqrt(2 \times pp(n)) < thresh$$

$$avoidance\ if\ thresh < TP(t=8) + sqrt(2 \times pp(n))$$

where engagement was chosen only if threat prediction is below threat threshold with > 95% confidence level. $thresh = 0.2$ was used for Figure 8.

**Reinforcement learning of salience prediction:** The above models propose that TS works together with a separate system that provides an approach drive. Risk assessment is performed when uncertainty of threat prediction is high, but not directly promoted by threat prediction. However, it is also possible that assessment is directly promoted by TS. Pearce and Hall proposed that attention to a specific stimulus is induced by prediction error of its outcome, which in turn promotes learning of the stimulus in the next trial (Pearce and Hall, 1980). Applying this idea, Gordon et al. modeled hierarchical reinforcement learning where prediction error promotes active sensing so that an agent is encouraged to learn what is

unexpected (Gordon and Ahissar, 2012; Gordon et al., 2014). The authors also combined it with the notion that too much novelty (prediction error) is fearful, causing retreat.

In the third model (Figure S5), we applied reinforcement learning to model prediction of prediction error, similar to hierarchical curiosity loops (Gordon and Ahissar, 2012; Gordon et al., 2014). The first order learner collects information of an object using a prediction error. To simplify, object information was modeled as a single dimension (e.g. size Φ), although multiple dimensions of object features are likely to be learned. The second agent models TS and learns prediction of object information gain (we will call "salience" here), which induces assessment, but also causes avoidance if the prediction is too high.

The object information V was updated only when an agent is at object ($t = 10$), following Rescorla-Wagner rule (Rescorla and Wagner, 1972).

$$\delta 1 = \Phi - V$$

$$V = V + \alpha \times \delta$$

The salience prediction at time $t$ is denoted as $SP(t)$, and is defined by,

$$SP(t) = E\left[\sum_{k=0}^{N-t} \gamma^k \cdot \ salience(t+k)\right]$$

similar to threat prediction in models 1 and 2. We assumed that a salience outcome occurred when the animal encountered a surprising feature of a novel object (i.e. $t = 10$).

$$salience(t = 10) = \delta 1, \quad threat(t \neq 10) = 0$$

Salience prediction, SP, was initialized to uniform small number 0.1 for the states approaching object.

$$SP(t) = 0.1 \quad for\ 0 < t < 10$$

Update rules for SP is the same as TP.

Behavior (risk assessment, engagement and avoidance) was chosen according to the salience prediction near the object, $TP(t = 8)$, compared to a threat threshold, *thresh, and an approach threshold, a_thresh*.

$$risk\ assessment\ if\ a\_thresh < SP(t = 8) < \ thresh$$

$$engagement\ if\ \ SP(t = 8) < a\_thresh$$

$$avoidance\ if\ thresh\ < SP(t = 8)$$

*thresh* = 0.28, *a_thresh* = 0.05 was used for Figure S5.

## QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis was performed using custom software written in MATLAB (MathWorks, Natick, MA, USA, RRID:SCR_001622). All error bars in the figures are SEM. In boxplots, the edges of the boxes are the 25$^{th}$ and 75$^{th}$ percentiles, and the whiskers extend to the most extreme data points not considered outliers. The exact value of p and n are indicated in figure legends unless otherwise noted.

**Time-course of behaviors—**Time spent near object is defined as fraction of time when the nose or tail fell within a radius of 7cm (Euclidean distance) from the center of the object (Figure 1B, Figure 3A, Figure 4B). Fraction of time spent near object per day and per min in individual animals, and average of all animals (mean ± SEM, n=26 animals) per min are shown in Figure 1B. Time spent near object was significantly correlated across novelty days, but not between novelty and habituation days (R=−0.02, p=0.89, H1; R=0.29, p=0.13, H2; R=0.87, p=0.0000, N2; R=0.69, p=0.001, N3; R=0.66, p=0.0002, N4, Pearson's correlation coefficient with N1, n=26 animals, Figure 1B). Fraction of time spent near object per min in individual animals are shown in Figure 3A and Figure 4B. Cumulative probability of each group of mice spending certain amounts of time near object on the first day of novelty (N1) is shown (Figure 3A, Figure 4B). Mice spend less time near a novel object than familiar object (p=0.018, n=9 animals for each group, Kolmogorov-Smirnov (K-S) test, Figure 3A). Ablation mice spend more time near a novel object than sham mice (p=0.030, n=17 animals for each group, K-S test, Figure 4B).

An approach bout is defined as an event from the time when either the nose or tail enters an area within 7cm from the center of the object to the time when both nose and tail are no longer within the area. Approach frequency is defined as frequency of approach bouts per min (Figure 1C), and approach bout duration is defined as average duration of approach bouts in 1 min (Figure 1D). Both data of individual animals and average of all animals (mean ± SEM, n=26 animals) are shown.

Distance from object is defined as distance between either nose or tail and the center of the object (Figure 2B). Closest point to object is defined as the shortest distance from the nose or tail to the center of the object in each bout (Figure 2C).

Approach bouts were broken down into two types depending on whether the nose was closer to the object than the tail for the entire bout (approach with tail behind) or whether the tail was closer at some point (approach with tail exposure). Frequency of approach with tail behind or tail exposure per min were calculated (Figure 2D–E, Figure 3B, Figure 4C). Both data of individual animals and average of all animals (mean ± SEM, n=26 animals for Figure 2D–E, n=9 animals for Figure 3B, n=17 animals for Figure 4C) are shown. For violin plots, average frequency are subtracted with average frequency in habituation days in each animal. Frequency of approach with tail behind decreases over time (p=2.8×10$^{-11}$, t-test,

n=26 animals, beta coefficients of linear regression of frequency with time, Figure 2D). Frequency of approach with tail behind does not show significant linear change over time (p=0.20, t-test, n=26 animals, beta coefficients of linear regression of frequency with time, Figure 2E). In boxplot in Figure 3C, average frequency of approach with tail behind and approach with tail exposure on N1 for each animal are shown. Approach with tail behind on N1 is more frequent towards a novel object than an unexpected familiar object (p=0.0031, n=9 animals for each group, t-test), whereas approach with tail exposure on N1 is more frequent towards an unexpected familiar object than a novel object (p=0.0031, n=9 animals for each group, t-test). Frequency of approach with tail behind on N1 was not significantly different between sham and ablation animals (p=0.069, n=17 animals for each, t-test) and approach with tail exposure on N1 was significantly more frequent in ablation animals than sham animals (p=0.010, n=17 animals for each, t-test). The distribution shape of data points was not formally tested. In Figure 3D, fraction of animals with approach with tail behind towards novel or unexpected familiar objects in each approach bout were plotted (total 9 animals).

**Moseq analysis—**Figure 5B shows fraction of video frames where each syllable is used in total video frames around retreat (−1s to 1s) in all approach bouts in all mice of the same condition. Syllable usage in each approach bout is shown above each plot. Figure 5C left shows fraction of approach bouts during which each syllable is used in all approach bouts in all novel object group at each time point. Syllable frequency is defined as frequency of emergence of each syllable regardless of duration of the syllable in the whole session (25 min), at all retreat (−1s to 1s), at retreat (−1s to 1s) of approach with tail behind, or at retreat (−1s to 1s) of approach with tail exposure in each animal (Figure 5C right). Figure 5E top shows average frequency of syllable usage in each group at each time point (mean ± SEM, n=9 animals for novel object and unexpected familiar object groups, n=17 animals for sham and ablation group). Figure 5E boxplots show distribution of total syllable expression on N1 in each animal (novel object vs unexpected familiar object, $p=4.9\times10^{-4}$, syllable 79; $p=4.9\times10^{-4}$, syllable 14, n=9 animals for each; sham vs ablation, p=0.010, syllable 79; p=0.030, syllable 14, n=17 animals for each, K-S test). Expression of both syllables decreased over time (−0.10/min, $p=6.8\times10^{-15}$, F-statistic 9.0; syllable 79; −0.07/min, $p=2.0\times10^{-12}$, F-statistic 7.2, syllable 14, linear regression of frequency of syllable usage with time and animals in the novel object group, degree of freedom 215). Figure 5F left shows fraction of each syllable expression following syllable 79 expression in sham and ablation animals. Figure 5F boxplots show distribution of fraction of syllable 14 expression following syllable 79 expression in sham and ablation animals (p=0.72, n=17 animals for each, t-test). Distribution shape of data points was not formally tested.

**Fluorometry analysis—**Z-scores were calculated using an entire recording session. Retreat start was defined as the time point when the animal's nose was closest to the object within an approach bout. Only one retreat start was detected in each approach bout to avoid using multiple time points close each other. Approach start was defined as the time point when the distance between the animal's nose and the object started decreasing before each retreat start. Retreat end was defined as the time point when the distance between the animal's nose and the object started decreasing after each retreat start. Responses aligned at

a behavioral event were calculated by subtracting the average baseline activity (−3s to −1s before the event) from the average activity of the target window (0–1s after the event). To show overall activity patterns (Figure 6A), the average activity (−3s to −1s before approach start) was used as baseline. Figure 6A bottom shows average dopamine sensor signals in all animals (mean ± SEM, n=15 animals). Figure 6B shows average dopamine sensor signals on N1 aligned to time of retreat in each animal.

Figure 6C plots average dopamine sensor signals of each animal against time spent near the object, frequency of approach with tail exposure, or time of the first approach with tail exposure in session. Dopamine sensor signals negatively correlate with time spent near the object ($R=-0.72$, $p=0.0022$), negatively correlate with frequency of approach with tail exposure ($R=-0.71$, $p=0.0028$), and positively correlate with time of the first approach with tail exposure in session ($R=0.80$, $p=3.2\times10^{-4}$) (Pearson's correlation coefficient, n=15 animals). First approach with tail exposure for mice that never showed approach with tail exposure (3 animals) was set to 25min, the last time point.

Figure 6D shows time-course of dopamine sensor signals across approach bouts ("trials") and time-course aligned to the first approach with tail exposure for each animal (total 15 animals). Figure 5E shows average dopamine sensor signals in mice that never showed approach with tail exposure (mean ± SEM, n=3 animals) and in mice that showed approach with tail exposure (mean ± SEM, n=12 animals). Approach bouts in animals with approach with tail exposure were divided into phase 1 and phase 2 by time of first approach with tail exposure. On average, dopamine response at retreat (0 to 1s) was higher in phase 1 than in phase 2 ($p=0.0059$, n=12 animals, paired t-test). Figure 6F shows dopamine sensor signals during phase 2 in mice that express approach with tail exposure (mean ± SEM, n=12 animals). On average, dopamine responses at retreat (0 to 1s) were similar between approach types in phase 2 ($p=0.90$, n=12 animals, paired t-test). These numbers are indicated in the main text. The distribution shape of data points was not formally tested.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENTS

We thank Adam Lowet, Malcolm Campbell, and all lab members for discussion and feedback. This work was supported by the NIH BRAIN Initiative (U19NS113201, NU, SRD; and R01NS108740, NU), National Institute of Mental Health (R01MH125162, MW-U), Simons Collaboration on the Global Brain (NU, SRD), Bipolar Disorder Seed Grant Program (NU), Japan Society for the Promotion of Science (IT-K), the Harvard Molecules, Cells, and Organisms Program training grant (KA), and a Career Award at the Scientific Interface from the Burroughs Wellcome Fund (JM). We thank the Harvard Center for Biological Imaging, Center for Brain Science Fabrication Lab, and the Physics/SEAS Machine Shop for technical support.

## REFERENCES

Baron-Cohen S, Knickmeyer RC, and Belmonte MK (2005). Sex differences in the brain: Implications for explaining autism. Science 310, 819–823. 10.1126/science.1115455. [PubMed: 16272115]

Barto A, Mirolli M, and Baldassarre G (2013). Novelty or Surprise? Frontiers in Psychology 4.

Blanchard DC, Blanchard RJ, and Rodgers RJ (1991). Risk Assessment and Animal Models of Anxiety. In Animal Models in Psychopharmacology, Olivier B, Mos J, and Slangen JL, eds. (Basel: Birkhäuser), pp. 117–134.

Bromberg-Martin ES, and Hikosaka O (2009). Midbrain Dopamine Neurons Signal Preference for Advance Information about Upcoming Rewards. Neuron 63, 119–126. 10.1016/j.neuron.2009.06.009. [PubMed: 19607797]

Cohen JY, Haesler S, Vong L, Lowell BB, and Uchida N (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature 482, 85–88. 10.1038/nature10754. [PubMed: 22258508]

Colas C, Fournier P, Chetouani M, Sigaud O, and Oudeyer P-Y (2019). CURIOUS: Intrinsically Motivated Modular Multi-Goal Reinforcement Learning. In Proceedings of the 36th International Conference on Machine Learning, (PMLR), pp. 1331–1340.

Corey DT (1978). The determinants of exploration and neophobia. Neuroscience & Biobehavioral Reviews 2, 235–253. 10.1016/0149-7634(78)90033-7.

Cox J, and Witten IB (2019). Striatal circuits for reward learning and decision-making. Nat Rev Neurosci 20, 482–494. 10.1038/s41583-019-0189-2. [PubMed: 31171839]

Dabney W, Kurth-Nelson Z, Uchida N, Starkweather CK, Hassabis D, Munos R, and Botvinick M (2020). A distributional code for value in dopamine-based reinforcement learning. Nature 577, 671–675. 10.1038/s41586-019-1924-6. [PubMed: 31942076]

Dai B, Sun F, Kuang A, Li Y, and Lin D (2021). Dopamine release in nucleus accumbens core during social behaviors in mice. 2021.06.22.449478. 10.1101/2021.06.22.449478.

Dayan P (1993). Improving Generalization for Temporal Difference Learning: The Successor Representation. Neural Computation 5, 613–624. 10.1162/neco.1993.5.4.613.

Dayan P (2021). “Liking” as a First Draft of the Affective Future. 10.31234/osf.io/g7zfq.

Eshel N, Tian J, and Uchida N (2013). Opening the black box: dopamine, predictions, and learning. Trends in Cognitive Sciences 17, 430–431. 10.1016/j.tics.2013.06.010. [PubMed: 23830895]

Fanselow MS (1994). Neural organization of the defensive behavior system responsible for fear. Psychon Bull Rev 1, 429–438. 10.3758/BF03210947. [PubMed: 24203551]

Fernandes AB, Alves da Silva J, Almeida J, Cui G, Gerfen CR, Costa RM, and Oliveira-Maia AJ (2020). Postingestive Modulation of Food Seeking Depends on Vagus-Mediated Dopamine Neuron Activity. Neuron 106, 778–788.e6. 10.1016/j.neuron.2020.03.009. [PubMed: 32259476]

Glimcher PW (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. Proceedings of the National Academy of Sciences 108, 15647–15654. 10.1073/pnas.1014269108.

Gordon G, and Ahissar E (2012). Hierarchical curiosity loops and active sensing. Neural Networks 32, 119–129. 10.1016/j.neunet.2012.02.024. [PubMed: 22386787]

Gordon G, Fonio E, and Ahissar E (2014). Emergent exploration via novelty management. J Neurosci 34, 12646–12661. 10.1523/JNEUROSCI.1872-14.2014. [PubMed: 25232104]

Gottlieb J, and Oudeyer P-Y (2018). Towards a neuroscience of active sampling and curiosity. Nat Rev Neurosci 19, 758–770. 10.1038/s41583-018-0078-0. [PubMed: 30397322]

Gunaydin LA, Grosenick L, Finkelstein JC, Kauvar IV, Fenno LE, Adhikari A, Lammel S, Mirzabekov JJ, Airan RD, Zalocusky KA, et al. (2014). Natural Neural Projection Dynamics Underlying Social Behavior. Cell 157, 1535–1551. 10.1016/j.cell.2014.05.017. [PubMed: 24949967]

Halliday MS (1966). Exploration and fear in the rat. Symp. Zool. Soc. London 18, 45–59.

Han W, Tellez LA, Perkins MH, Perez IO, Qu T, Ferreira J, Ferreira TL, Quinn D, Liu Z-W, Gao X-B, et al. (2018). A Neural Circuit for Gut-Induced Reward. Cell 175, 665–678.e23. 10.1016/j.cell.2018.08.049. [PubMed: 30245012]

He K, Zhang X, Ren S, and Sun J (2016). Deep Residual Learning for Image Recognition. pp. 770–778.

Hirshfeld-Becker DR, Micco JA, Wang CH, and Henin A (2014). Behavioral inhibition: A discrete precursor to social anxiety disorder? The Wiley Blackwell Handbook of Social Anxiety Disorder 133–158. 10.1002/9781118653920.ch7.

Hogan JA (1965). An Experimental Study of Conflict and Fear: an Analysis of Behavior of Young Chicks Toward a Mealworm. Part I. the Behavior of Chicks Which Do Not Eat the Mealworm. Behaviour 25, 45–96. 10.1163/156853965X00110. [PubMed: 5824947]

Horvitz JC, Stewart T, and Jacobs BL (1997). Burst activity of ventral tegmental dopamine neurons is elicited by sensory stimuli in the awake cat. Brain Research 759, 251–258. 10.1016/S0006-8993(97)00265-5. [PubMed: 9221945]

Hughes RN (1997). Intrinsic exploration in animals: motives and measurement. Behavioural Processes 41, 213–226. 10.1016/S0376-6357(97)00055-7. [PubMed: 24896854]

Insafutdinov E, Pishchulin L, Andres B, Andriluka M, and Schiele B (2016). DeeperCut: A Deeper, Stronger, and Faster Multi-person Pose Estimation Model. In Computer Vision – ECCV 2016, Leibe B, Matas J, Sebe N, and Welling M, eds. (Cham: Springer International Publishing), pp. 34–50.

Jaegle A, Mehrpour V, and Rust N (2019). Visual novelty, curiosity, and intrinsic reward in machine learning and the brain. Current Opinion in Neurobiology 58, 167–174. 10.1016/j.conb.2019.08.004. [PubMed: 31614282]

Jiujias M, Kelley E, and Hall L (2017). Restricted, Repetitive Behaviors in Autism Spectrum Disorder and Obsessive–Compulsive Disorder: A Comparative Review. Child Psychiatry and Human Development 48, 944–959. 10.1007/s10578-017-0717-0. [PubMed: 28281020]

Kagan J, Reznick JS, Clarke C, Snidman N, and Garcia-coll C (1984). Behavioral Inhibition to the Unfamiliar. Child Development 55, 2212–2225.

Kakade S, and Dayan P (2002). Dopamine: Generalization and bonuses. Neural Networks 15, 549–559. 10.1016/S0893-6080(02)00048-5. [PubMed: 12371511]

Kaplan F, and Oudeyer P-Y (2007). In search of the neural circuits of intrinsic motivation. Frontiers in Neuroscience 1.

Kidd C, and Hayden BY (2015). The Psychology and Neuroscience of Curiosity. Neuron 88, 449–460. 10.1016/j.neuron.2015.09.010. [PubMed: 26539887]

Kim HF, and Hikosaka O (2013). Distinct Basal Ganglia Circuits Controlling Behaviors Guided by Flexible and Stable Values. Neuron 79, 1001–1010. 10.1016/j.neuron.2013.06.044. [PubMed: 23954031]

Kim HF, Ghazizadeh A, and Hikosaka O (2015). Dopamine Neurons Encoding Long-Term Memory of Object Value for Habitual Behavior. Cell 163, 1165–1175. 10.1016/j.cell.2015.10.063. [PubMed: 26590420]

Kumaran D, and Maguire EA (2007). Which computational mechanisms operate in the hippocampus during novelty detection? Hippocampus 17, 735–748. 10.1002/hipo.20326. [PubMed: 17598148]

Lak A, Stauffer WR, and Schultz W (2016). Dopamine neurons learn relative chosen value from probabilistic rewards. ELife 5, e18044. 10.7554/eLife.18044. [PubMed: 27787196]

Lerner TN, Shilyansky C, Davidson TJ, Evans KE, Beier KT, Zalocusky KA, Crow AK, Malenka RC, Luo L, Tomer R, et al. (2015). Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. Cell 162, 635–647. 10.1016/j.cell.2015.07.014. [PubMed: 26232229]

Lester D (1967). Sex Differences in Exploration: Toward a Theory of Exploration. Psychol Rec 17, 55–62. 10.1007/BF03393689.

Ljungberg T, Apicella P, and Schultz W (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. Journal of Neurophysiology 67, 145–163. 10.1152/jn.1992.67.1.145. [PubMed: 1552316]

Lowet AS, Zheng Q, Matias S, Drugowitsch J, and Uchida N (2020). Distributional Reinforcement Learning in the Brain. Trends in Neurosciences 43, 980–997. 10.1016/j.tins.2020.09.004. [PubMed: 33092893]

Marder E, and Goaillard J-M (2006). Variability, compensation and homeostasis in neuron and network function. Nat Rev Neurosci 7, 563–574. 10.1038/nrn1949. [PubMed: 16791145]

Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, and Bethge M (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat Neurosci 21, 1281–1289. 10.1038/s41593-018-0209-y. [PubMed: 30127430]

Menegas W, Babayan BM, Uchida N, and Watabe-Uchida M (2017). Opposite initialization to novel cues in dopamine signaling in ventral and posterior striatum in mice. ELife 6, e21886. 10.7554/eLife.21886. [PubMed: 28054919]

Menegas W, Akiti K, Amo R, Uchida N, and Watabe-Uchida M (2018). Dopamine neurons projecting to the posterior striatum reinforce avoidance of threatening stimuli. Nat Neurosci 21, 1421–1430. 10.1038/s41593-018-0222-1. [PubMed: 30177795]

Montague PR, Dayan P, and Sejnowski TJ (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J. Neurosci 16, 1936–1947. 10.1523/JNEUROSCI.16-05-01936.1996. [PubMed: 8774460]

Montgomery KC (1955). The relation between fear induced by novel stimulation and exploratory drive. Journal of Comparative and Physiological Psychology 48, 254–260. 10.1037/h0043788. [PubMed: 13252152]

Morrens J, Aydin Ç, Janse van Rensburg A, Esquivelzeta Rabell J, and Haesler S (2020). Cue-Evoked Dopamine Promotes Conditioned Responding during Learning. Neuron 106, 142–153.e7. 10.1016/j.neuron.2020.01.012. [PubMed: 32027824]

Ng AY, Harada D, and Russell S (1999). Policy invariance under reward transformations: Theory and application to reward shaping. In Icml, pp. 278–287.

Ogasawara T, Sogukpinar F, Zhang K, Feng Y-Y, Pai J, Jezzini A, and Monosov IE (2022). A primate temporal cortex–zona incerta pathway for novelty seeking. Nat Neurosci 25, 50–60. 10.1038/s41593-021-00950-1. [PubMed: 34903880]

Orefice LLL, Zimmerman ALL, Chirila AMM, Sleboda SJJ, Head JPP, and Ginty DDD (2016). Peripheral Mechanosensory Neuron Dysfunction Underlies Tactile and Behavioral Deficits in Mouse Models of ASDs. Cell 166, 299–313. 10.1016/j.cell.2016.05.033. [PubMed: 27293187]

Oudeyer P-Y, Kaplan F, and Hafner VV (2007). Intrinsic Motivation Systems for Autonomous Mental Development. IEEE Transactions on Evolutionary Computation 11, 265–286. 10.1109/TEVC.2006.890271.

Oudeyer P-Y, Gottlieb J, and Lopes M (2016). Chapter 11 - Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. In Progress in Brain Research, Studer B, and Knecht S, eds. (Elsevier), pp. 257–284.

Parker NF, Cameron CM, Taliaferro JP, Lee J, Choi JY, Davidson TJ, Daw ND, and Witten IB (2016). Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. Nat Neurosci 19, 845–854. 10.1038/nn.4287. [PubMed: 27110917]

Paxinos G, and Franklin KBJ (2019). Paxinos and Franklin's the Mouse Brain in Stereotaxic Coordinates (Academic Press).

Pearce JM, and Hall G (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological Review 87, 532–552. 10.1037/0033-295X.87.6.532. [PubMed: 7443916]

Ranganath C, and Rainer G (2003). Neural mechanisms for detecting and remembering novel events. Nat Rev Neurosci 4, 193–202. 10.1038/nrn1052. [PubMed: 12612632]

Rebec GV, Christensen JRC, Guerra C, and Bardo MT (1997). Regional and temporal differences in real-time dopamine efflux in the nucleus accumbens during free-choice novelty. Brain Research 776, 61–67. 10.1016/S0006-8993(97)01004-4. [PubMed: 9439796]

Rescorla R, and Wagner A (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Classical Conditioning II: Current Research and Theory, p.

Schlegel M, Jacobsen A, Abbas Z, Patterson A, White A, and White M (2021). General Value Function Networks. Journal of Artificial Intelligence Research 70, 497–543. 10.1613/jair.1.12105.

Schultz W (1998). Predictive reward signal of dopamine neurons. Journal of Neurophysiology 80, 1–27. [PubMed: 9658025]

Schultz W (2015). Neuronal Reward and Decision Signals: From Theories to Data. Physiological Reviews 95, 853–951. 10.1152/physrev.00023.2014. [PubMed: 26109341]

Schultz W (2016). Dopamine reward prediction-error signalling: a two-component response. Nat Rev Neurosci 17, 183–195. 10.1038/nrn.2015.26. [PubMed: 26865020]

Schultz W, Dayan P, and Montague PR (1997). A neural substrate of prediction and reward. Science 275, 1593–1599. 10.1126/science.275.5306.1593. [PubMed: 9054347]

Simsek M, Czylwik A, Galindo-Serrano A, and Giupponi L (2011). Improved decentralized Q-learning algorithm for interference reduction in LTE-femtocells. In 2011 Wireless Advanced, pp. 138–143.

Singh S, Lewis RL, Barto AG, and Sorg J (2010). Intrinsically Motivated Reinforcement Learning: An Evolutionary Perspective. IEEE Transactions on Autonomous Mental Development 2, 70–82. 10.1109/TAMD.2010.2051031.

Skinner BF (1975). The shaping of phylogenic behavior. Journal of the Experimental Analysis of Behavior 24, 117–120. [PubMed: 16811859]

Stout A, Konidaris GD, and Barto AG (2005). Intrinsically Motivated Reinforcement Learning: A Promising Framework for Developmental Robot Learning (MASSACHUSETTS UNIV AMHERST DEPT OF COMPUTER SCIENCE).

Sun F, Zhou J, Dai B, Qian T, Zeng J, Li X, Zhuo Y, Zhang Y, Wang Y, Qian C, et al. (2020). Next-generation GRAB sensors for monitoring dopaminergic activity in vivo. Nat Methods 17, 1156–1166. 10.1038/s41592-020-00981-9. [PubMed: 33087905]

Sutton RS, and Barto AG (1987). A temporal-difference model of classical conditioning. In Proceedings of the Ninth Annual Conference of the Cognitive Science Society, (Seattle, WA), pp. 355–378.

Sutton RS, and Barto AG (1990). Time-Derivative Models of Pavlovian Reinforcement. Learning and Computational Neuroscience: Foundations of Adaptive Networks 497–537. 10.1111/j.1748-1716.1960.tb01900.x.

Sutton RS, and Barto AG (2018). Reinforcement Learning, second edition: An Introduction (MIT Press).

Tellez LA, Han W, Zhang X, Ferreira TL, Perez IO, Shammah-Lagnado SJ, van den Pol AN, and de Araujo IE (2016). Separate circuitries encode the hedonic and nutritional values of sugar. Nat Neurosci 19, 465–470. 10.1038/nn.4224. [PubMed: 26807950]

Thiele SL, Warre R, and Nash JE (2012). Development of a Unilaterally-lesioned 6-OHDA Mouse Model of Parkinson's Disease. JoVE (Journal of Visualized Experiments) e3234. 10.3791/3234.

Thorpe WH (1956). Learning and instinct in animals (Cambridge, MA, US: Harvard University Press).

Tsutsui-Kimura I, Matsumoto H, Akiti K, Yamada MM, Uchida N, and Watabe-Uchida M (2020). Distinct temporal difference error signals in dopamine axons in three regions of the striatum in a decision-making task. ELife 9, e62390. 10.7554/eLife.62390. [PubMed: 33345774]

Watabe-Uchida M, and Uchida N (2018). Multiple Dopamine Systems: Weal and Woe of Dopamine. Cold Spring Harb Symp Quant Biol 83, 83–95. 10.1101/sqb.2018.83.037648. [PubMed: 30787046]

Wiewiora E (2003). Potential-Based Shaping and Q-Value Initialization are Equivalent. Journal of Artificial Intelligence Research 19, 205–208. 10.1613/jair.1190.

Wiltschko AB, Johnson MJ, Iurilli G, Peterson RE, Katon JM, Pashkovski SL, Abraira VE, Adams RP, and Datta SR (2015). Mapping Sub-Second Structure in Mouse Behavior. Neuron 88, 1121–1135. 10.1016/j.neuron.2015.11.031. [PubMed: 26687221]

Xu HA, Modirshanechi A, Lehmann MP, Gerstner W, and Herzog MH (2021). Novelty is not surprise: Human exploratory and adaptive behavior in sequential decisionmaking.

## Highlights

- Novelty-induced behaviors are analyzed using modern machine-learning methods
- Novelty induces risk assessment which develops into engagement or avoidance
- Dopamine in the tail of striatum correlates with individual behavioral variability
- Reinforcement learning with shaping bonus and uncertainty explains the data

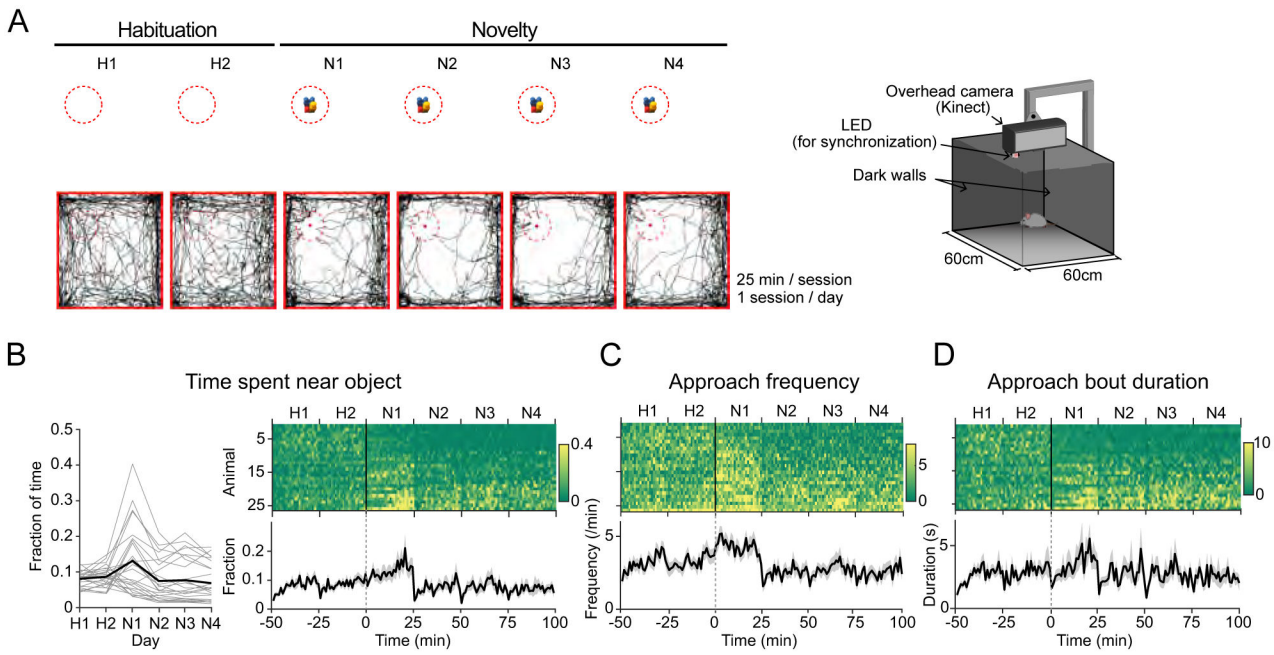


**Figure 1. Diversity of novelty behavior is captured in open arena**

**A.** Trajectory of nose from an example animal in the first 10 minutes of each session. **B.** Time spent within object area (7cm radius). Left thick black, average value across mice. Right bottom, mean ± SEM. Time spent near object was significantly correlated across novelty days, but not between novelty and habituation days (R=−0.02, p=0.89, H1; R=0.29, p=0.13, H2; R=0.87, p=0.0000, N2; R=0.69, p=0.001, N3; R=0.66, p=0.0002, N4, Pearson's correlation coefficient with N1, n=26 animals). **C.** Frequency of approaches. **D.** Duration of approach bouts.

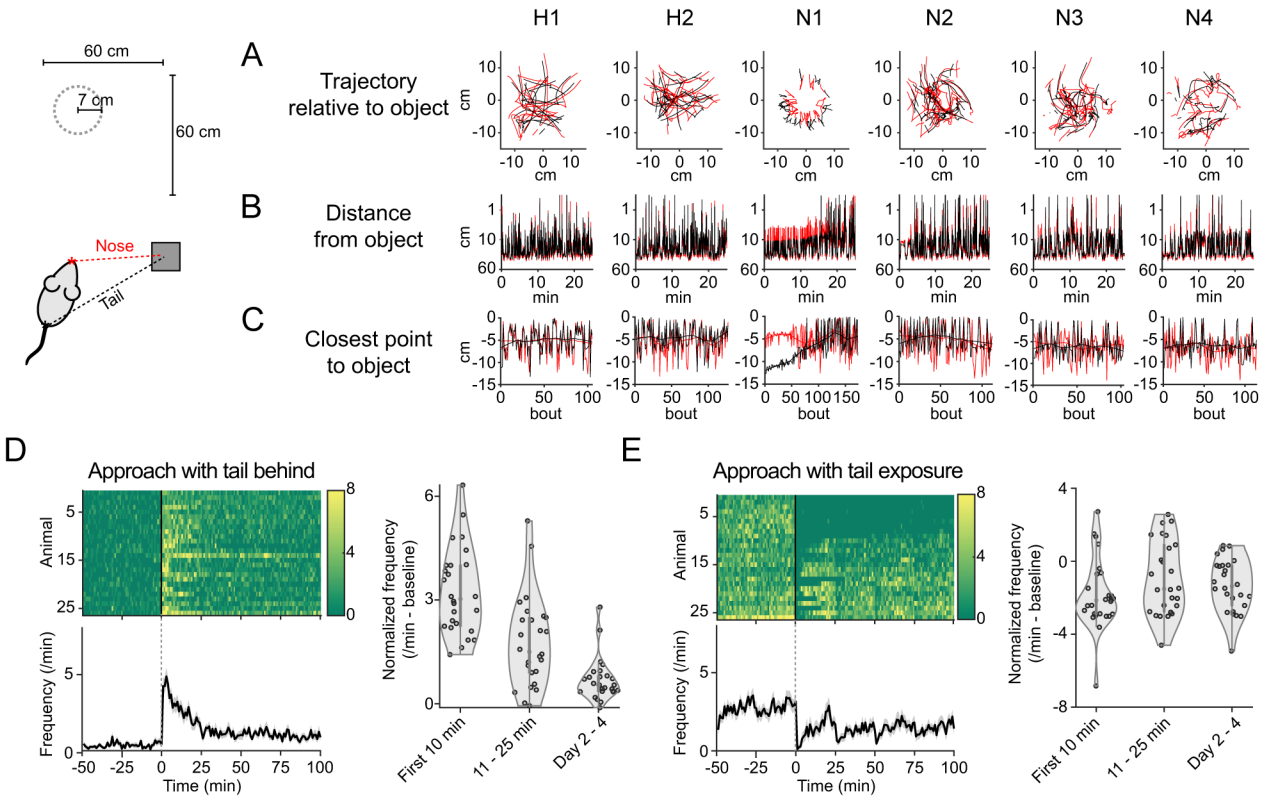


**Figure 2. Stereotypic behavioral response to novelty.**

**A.** Trajectory of nose or tail (in red and black, respectively) from an example mouse in the first 20 bouts of each session. **B.** Nose and tail position relative to object in an example animal. **C.** The closest position to object within each bout for nose and tail in an example animal. **D.** Frequency of approach bout with tail behind. Bottom, mean ± SEM. Right, average frequency normalized with baseline on habituation for each mouse. Tail behind approach frequency decreases over time ($p=2.8\times10^{-11}$, t-test, n=26 animals, beta coefficients of linear regression of frequency with time). **E.** Fraction of tail exposure.

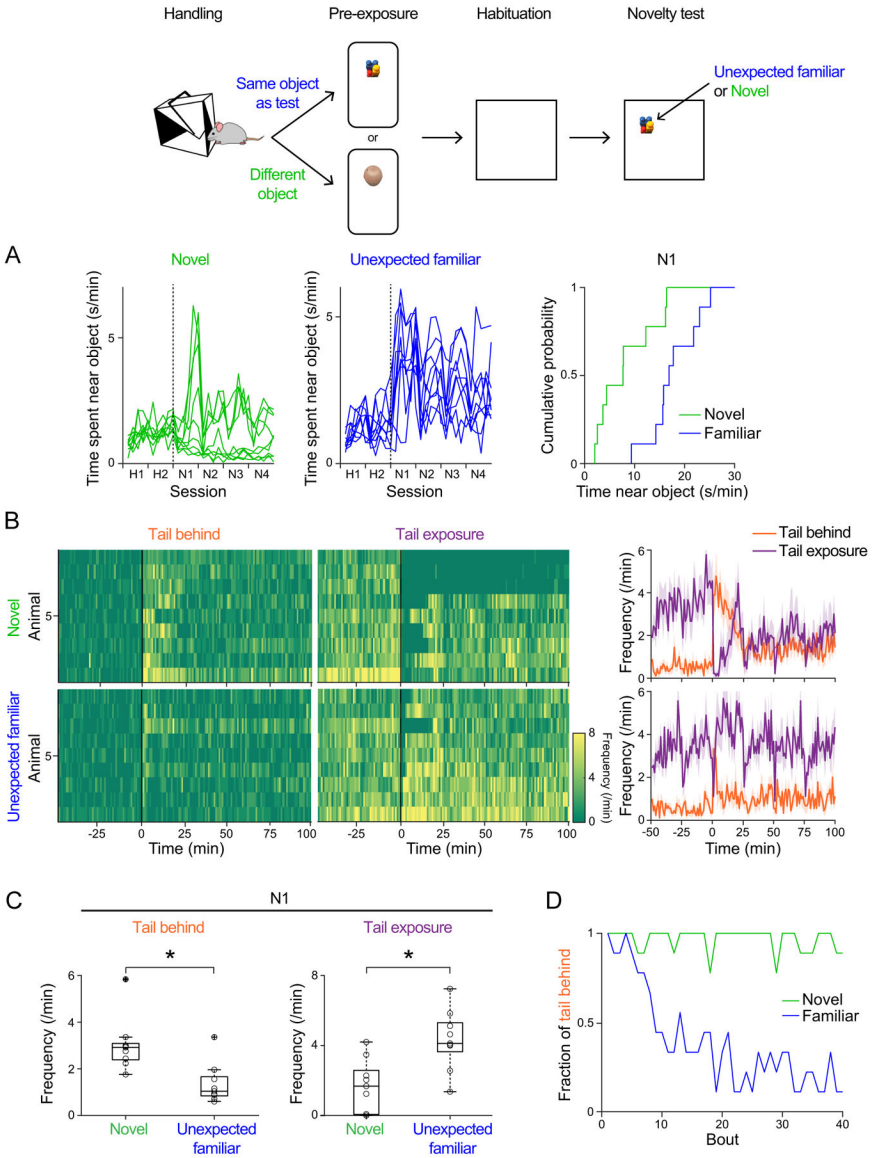


**Figure 3. Suppression of post-assessment engagement with stimulus novelty.**

**A.** Time spent near an object. Right, cumulative probability on N1. Mice spend less time near a novel object (p=0.018, n=9 animals for each group, Kolmogorov-Smirnov (K-S) test). **B.** Frequency of each approach type. Right, mean ± SEM. **C.** Average frequency of approaches on N1 for each mouse. Approach with tail behind is more frequent towards novel objects (p=0.0031), whereas approach with tail exposure is more frequent towards unexpected familiar objects (p=0.0031, n=9 animals for each group, t-test). **D.** Fraction of animals with approach with tail behind in each approach bout.

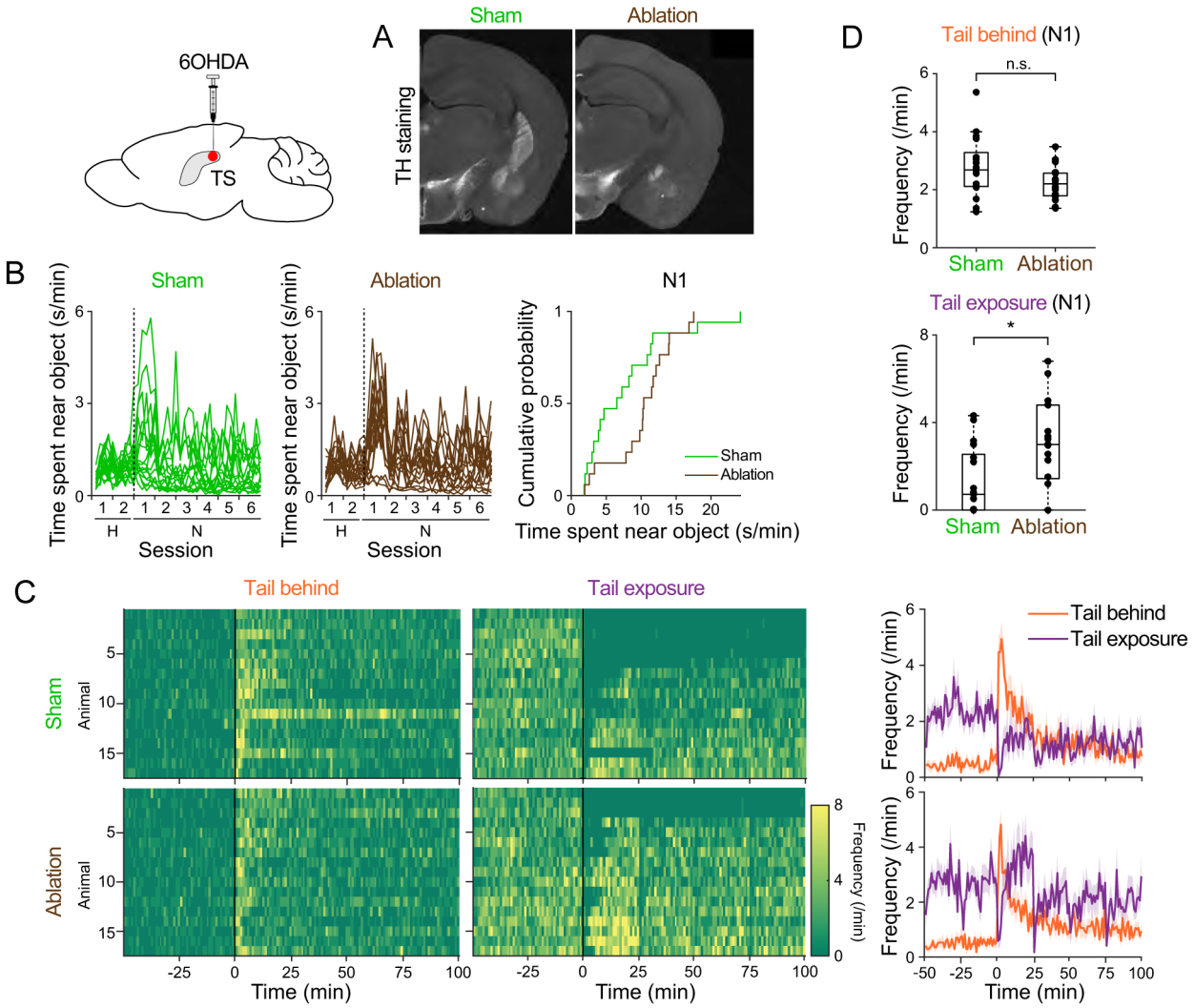


**Figure 4. Ablation of TS-projecting dopamine neurons promotes post-assessment engagement.**

**A.** Coronal sections (bregma −1.5mm) from sham (left) and ablation (right) animals. Dopamine axons were labeled with anti-tyrosine hydroxylase (TH) antibody. BLA, basolateral amygdala; CeA, central amygdala. **B.** Time spent near object. Right, cumulative probability on N1. Ablation vs sham, p=0.030 (K-S test). **C.** Frequency of each approach type bouts. Right, mean ± SEM. **D.** Average frequency of approach with tail behind (left; p=0.069, t-test) and approach with tail exposure (right, p=0.010, t-test) on N1. n=17 animals for each group. See also Figure S1 and Figure S2.

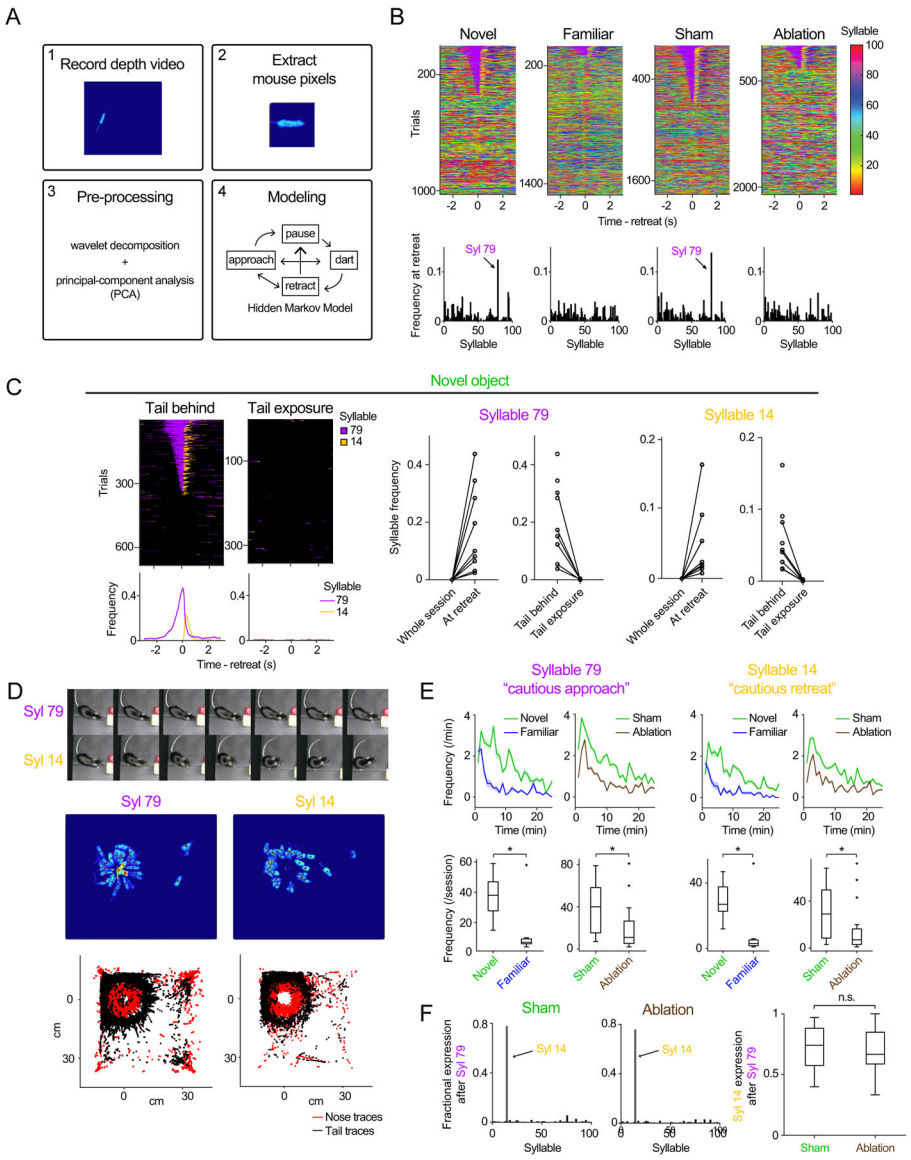


**Figure 5. Behavioral segmentation of novelty responses using MoSeq**

**A.** MoSeq workflow. **B.** Top, syllable usage across all approach bouts on N1 in all mice. Bottom, fraction of syllable usage at retreat (−1s to 1s). **C.** Syllable usage in novel object group. **D.** Top, example image series and superimposed images (full videos in Video S1 and S2). Bottom, spatial expression. **E.** Syllable usage in each group. Top, time-course (mean ± SEM). Bottom, total syllable expression (novel object vs unexpected familiar object, $p=4.9\times10^{-4}$, syllable 79; $p=4.9\times10^{-4}$, syllable 14, n=9 animals for each; sham vs ablation, p=0.010, syllable 79; p=0.030, syllable 14, n=17 animals for each, K-S test). Expression of both syllables decreased over time (−0.10/min, $p=6.8\times10^{-15}$, F-statistic 9.0; syllable 79; −0.07/min, $p=2.0\times10^{-12}$, F-statistic 7.2, syllable 14, linear regression with time and animals in the novel object group, degree of freedom 215). **F.** Left, fractional expression of each syllable after syllable 79. Right, fraction of syllable 14 expression following syllable 79 expression (p=0.72, n=17 animals for each, t-test). See also Figure S3.

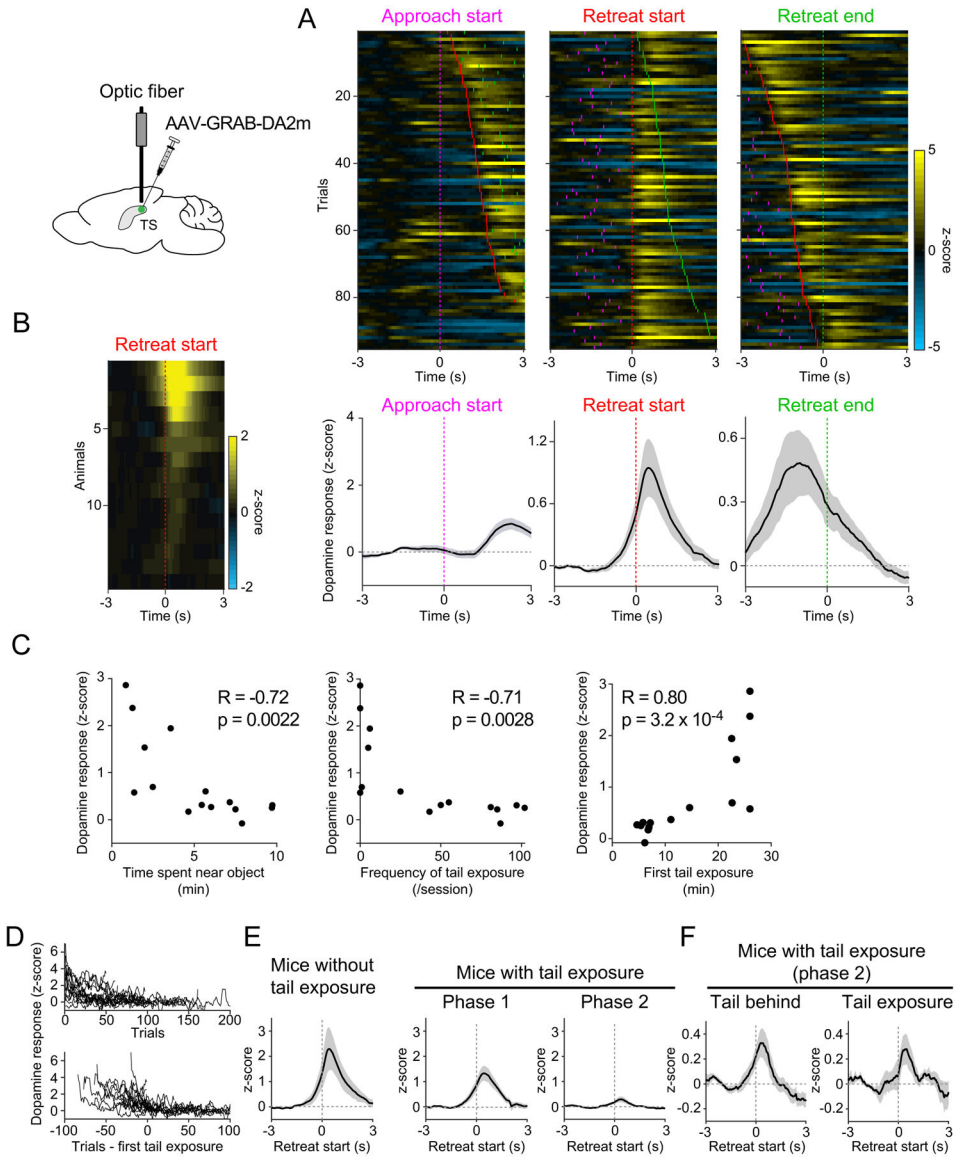


**Figure 6. Individual variability in behavior correlates with dopamine in TS.**

**A.** Dopamine signals in each trial in an example animal (top) and mean ± SEM (bottom, n=15 animals). Tick marks, approach start (cyan), retreat start (red), and retreat end (green). **B.** Average dopamine signals on N1 in each animal. **C**. Average dopamine signals of each animal plotted against behavioral measurements and Pearson's correlation coefficient, n=15 animals. First tail exposure for mice that never showed tail exposure (3 animals) was set to 25min. **D.** Time-course of dopamine signals across trials for each animal (top) or aligned to the first tail exposure (bottom). **E.** Dopamine signals in mice that never showed approach with tail exposure (left, n=3 animals) and in other mice (right, n=12 animals). mean ± SEM. **F.** Dopamine signals during phase 2. mean ± SEM, n=12 animals. See also Figure S4.

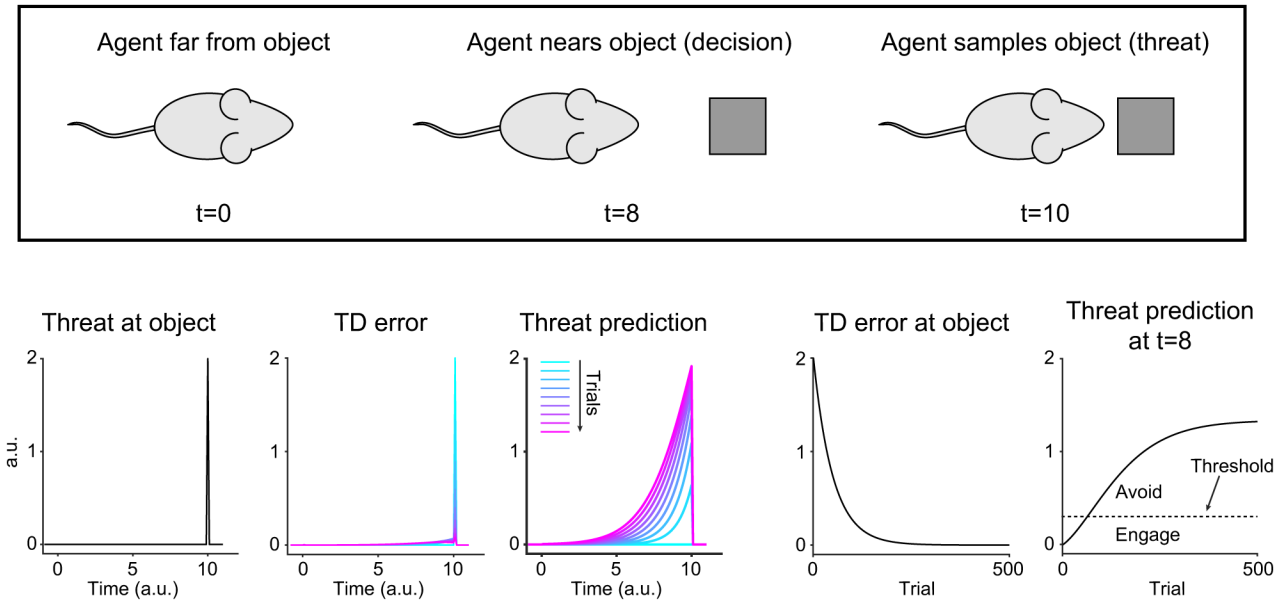


**Figure 7. Basic reinforcement learning model with constant threat.**

The time-course of variables within each trial (left) and over trials (right). Color, Trial 1–161, every 20 trials.

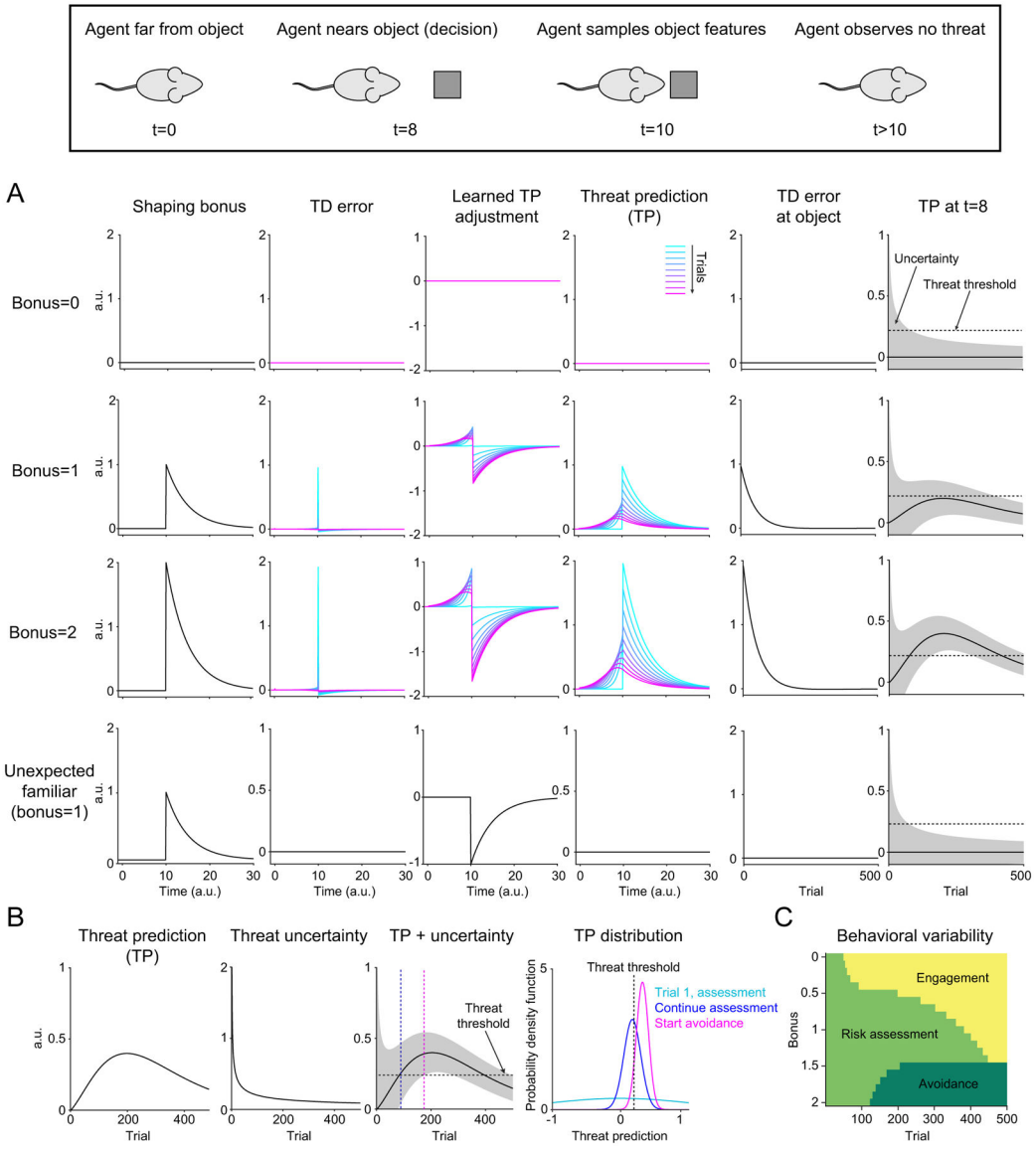


**Figure 8. Reinforcement learning model with shaping bonus and uncertainty.**

**A.** The time-course of variables within each trial (left) and over trials (right). Color, trial 1–321, every 40 trials. **B.** Components to determine behaviors. Left, threat prediction near object (t=8). Second from left, threat uncertainty near object. Third from left, threat prediction plotted together with threat uncertainty (shading). Black dotted line, threat threshold. Right, threat prediction distribution in example trials (trial 1 and trials shown with blue and cyan dotted line in third from left). **C.** Development of behaviors based on different degrees of shaping bonus. See also Figure S5.

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| Anti-tyrosine hydroxylase (TH) | EMD Millipore | RRID: AB_390204 |
| Alexa Fluor 594 goat anti-rabbit secondary antibody | Invitrogen | RRID: AB_2534079 |
| Bacterial and virus strains | | |
| AAV9-hSyn-GRABDA2m | Vigene Biosciences | YL002009-AV9-PUB |
| AAV5-CAG-tdTomato | UNC Vector Core | AAV In Stock<br>Vectors: Ed Boyden |
| AAV5-CAG-GFP | UNC Vector Core | AAV In Stock<br>Vectors: Ed Boyden |
| Chemicals, peptides, and recombinant proteins | | |
| 6-Hydroxydopamine hydrochloride | Sigma-Aldrich | H4381 |
| Deposited data | | |
| Matlab codes | This paper | GitHub (https://github.com/ckakiti/Novelty_paper_2021) |
| Video tracking and dopamine fluorometry data | This paper | Dryad (doi:10.5061/dryad.41ns1rnh2) |
| Experimental models: Organisms/strains | | |
| Mouse: C57BL/6J | Jackson Laboratory | RRID: IMSR_JAX:000664 |
| Software and algorithms | | |
| DeepLabCut | Mathis et al., 2018 | https://github.com/DeepLabCut/DeepLabCut |
| MoSeq | Wiltschko et al., 2015 | https://dattalab.github.io/moseq2-website/index.html |
| MATLAB | MathWorks | RRID:SCR_001622 |
| LabView | National Instruments | RRID:SCR_014325 |
| Other | | |
| Novelty object: LEGO | Mega Bloks | DCH63 |
| Novelty object: rubber toy | Kong Classic, M | https://www.kongcompany.com/kong-classic |
| Lighting in Novelty arena | Home Depot | https://www.homedepot.com/p/Westek-Indoor-Outdoor-6-ft-White-LED-Rope-Light-Kit-LROPE6W/312080910?source=shoppingads&lo |
| Camera in Novelty arena | Xbox | [discontinued] https://www.target.com/p/xbox-one-stand-alone-kinect-sensor/-/A-16504446?AFID=google_pla_df&CPNG=PLA_Electronics%20Shopping&LID=700000001170770pgs&adgroup=SC_Electronics&device=c&gclid= |
| Camera adaptor | Xbox | https://www.amazon.com/perseids-Adapter-Windows-Interactive-Development/dp/B07CWQK6XG/ref=sr_1_5?ie=UTF8&keywords=ine |

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Novelty arena: Frame | McMaster-Carr | https://www.mcmaster.com/47065T101 |
| Novelty arena: Walls/floor (outside) | McMaster-Carr | https://www.mcmaster.com/8505k744 |
| Novelty arena: Walls (inside) | McMaster-Carr | https://www.thorlabs.com/thorproduct.cfm?partnumber=BFP1 |
| Novelty arena: Floor (inside) | McMaster-Carr | https://www.walmart.com/ip/Five-Star-2-Pocket-Stay-Put-Plastic-Folder-Red-72109/310380845 |
| Repository with instructions for running code | This paper | https://github.com/ckakiti/Novelty_paper_2021 |