Research Article

The dopamine dip amplitude is a quantitative measure of disappointment

Yu Shikano[1]*, Sho Yagishita[2], Kenji F. Tanaka[1], Norio Takata[1]*

[1] Department of Neuropsychiatry, Keio University School of Medicine, 35 Shinanomachi, Shinjuku, Tokyo 160-8582, Japan

[2] Center for Disease Biology and Integrative Medicine, Faculty of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-0033, Japan

*For correspondence

Yu Shikano, Ph.D. Email: shikano_1-yu@yahoo.co.jp

Norio Takata, Ph.D. Email: takata.norio@keio.jp

Author ORCIDs

Yu Shikano : 0000-0001-8890-7691

Sho Yagishita : 0000-0003-2854-773X

Kenji F. Tanaka : 0000-0003-2511-0057

Norio Takata: 0000-0003-2754-8562

Abstract

Great expectations can lead to greater disappointment when the desired results are not achieved. Although it is well-documented that a better-than-expected reward is encoded quantitatively via midbrain dopaminergic (DA) activity, it has been less addressed experimentally whether disappointment, a worse-than-expected outcome is also expressed quantitatively by the DA signal. We show that the degree of disappointment is quantified by the magnitude of a transient decrease in the extracellular DA concentration (DA dip) in the ventral striatum of mice. We set up a lever press task on a fixed ratio (FR) schedule requiring five lever presses as an effort for a food reward (FR5). Mice occasionally checked the food magazine without a reward before completing the task. The percentage of this premature magazine entry (PME) increased as the number of lever presses approached five. This behavioral readout shows rising expectations with increasing proximity to task completion, leading to greater disappointment in the mice. Fiber photometry of the extracellular DA dynamics in the ventral striatum using a fluorescent protein (genetically encoded GPCR-activation-based-DA sensor: $GRAB_{DA2m}$) revealed that the amplitude of the DA dip following a PME was correlated with the percentage of the PME, demonstrating a monotonic relationship between the DA dip amplitude and the degree of disappointment. Computational modeling of the lever press task implementing temporal difference errors and state transitions replicated the positive correlation between the PME frequency and DA dip amplitude during improvement of the FR5 task. Taken together, these findings indicate that the degree of disappointment while making an effort toward a goal is represented monotonically by the magnitude of the DA dip in the ventral striatum, which may guide behavioral adjustment.

Keywords:

Dopamine, operant conditioning, striatum, GRAB sensor, fiber photometry

Introduction

Learning an appropriate association between behavior and outcome is critical for animal survival. Reinforcement learning (RL) algorithms are proposed as neuronal implementations for such experience-based learning (***Schultz et al., 1997***). Such an algorithm updates the subjective value of a state or an action based on the difference between the expected reward and the actual reward obtained, which is called the reward prediction error (RPE). An increasing number of studies have demonstrated that a better-than-expected outcome (i.e., a positive RPE) is encoded quantitatively by spiking activity of midbrain dopaminergic (DA) neurons (***Bayer and Glimcher, 2005; Tobler et al., 2005***) and by extracellular DA levels in the ventral striatum (***Covey and Cheer, 2019; Heien et al., 2005***). In contrast, there are still ongoing debates regarding the neuronal representation of disappointment, a worse-than-expected outcome inducing a negative RPE (***Hart et al., 2014***). While positive and negative RPEs are treated in the same way in RL algorithms, it is not clear whether the neuronal implementations of negative RPEs are similar to those of positive RPEs. Indeed, the economic expected utility theory (***Neumann and Morgenstern, 1945***) and the psychological prospect theory (***Kahneman and Tversky, 1979***) claim that the subjective perceptions of positive and negative rewards are *a*symmetric. Additionally, a biological investigation revealed that a DA surge and decrease in positive and negative RPEs, respectively, evoke distinct biochemical responses in different downstream targets (***Iino et al., 2020***).

DA dynamics in the ventral striatum have been investigated using an electrophysiological measurement of midbrain DA neurons that send afferents to the ventral striatum. Whereas firing activity of the DA neurons can be observed with a high temporal resolution, DA release dynamics in the ventral striatum cannot be determined, as spiking-activity-independent DA release has been identified (***Mohebi et al., 2019***). Fast scan cyclic voltammetry (FSCV) has been a standard method for the direct detection of extracellular DA, extending our knowledge of DA dynamics during behavior (***Heien et al., 2005; Phillips et al., 2003; Robinson et al., 2003; Rodeberg et al., 2017***). Recently, a genetically encoded optical DA sensor, G protein-coupled receptor-activation-based DA ($GRAB_{DA2m}$) was reported (***Sun et al., 2018; Sun et al., 2020***). $GRAB_{DA2m}$ has a high sensitivity (EC_{50} values of 90 nM), high specificity (10–20-fold more selective for DA over structurally similar noradrenaline) and long-term stability (> 120 min), suited for addressing the debate on a DAergic representation of a negative RPE (***Hamid et al., 2016***).

In the present study, we examined the neuronal representation of disappointment using a DA signal in the ventral striatum. We developed a fiber photometry system equipped with a highly sensitive Gallium Arsenide Phosphide (GaAsP) photocathode and combined this system with the viral expression of $GRAB_{DA2m}$ proteins in the ventral striatum of mice.

We monitored the DA dynamics during a lever press task on a fixed ratio (FR) schedule. The mice had to press a lever five times to obtain a food reward (FR5) (***Yoshida et al., 2019***). The advantage of this simple operant task is that it brought about various degrees of disappointment to the mice 1) using a fixed reward size (always a single pellet) after the completion of a FR5 task, and 2) during almost identical behavior patterns in anticipation of a reward. These characteristics contrast with other popular tasks that use variable reward sizes (e.g., the amount of juice) or forces a behavioral choice (e.g., left or right) for modulating the magnitude of expectation—either of which may introduce extra DA dynamics unrelated to disappointment. We found that the amplitude of a transient decrease in the extracellular DA (DA dip) reflected the magnitude of disappointment that was estimated based on behavior. An RL model using the temporal-difference (TD) algorithm (***Amo et al., 2020; Schultz et al., 1997***) supported that the modulation of the DA dip amplitude according to the degree of disappointment was important for the mice to establish adaptive behavior in the FR5 task. Taken together, we propose that the degree of disappointment is represented quantitatively by the magnitude of the DA dip, which plays a role in behavior adjustment.

Results

Results 1: Completion Latency Decreases with Proficiency in the FR5 Task

We first trained food-restricted mice to press a lever to earn a food pellet. An operant task on a fixed ratio (FR) 1 schedule was used to establish the response (lever press)–outcome (a food pellet delivery) relationship (***Fig. 1A***). In the task, the trial started with a sound cue 5 s in duration, followed by the presentation of two levers (lever extension). Only the left lever was reinforced; a single press of the reinforced lever resulted in a food reward delivery and the retraction of both levers. Then, the mice nose-poked the food magazine to collect a food pellet (magazine entry, ME). The mice engaged in the series of these events to complete a trial (completed trial). Then, 30 seconds of an inter-trial interval (ITI) was added, followed by the next trial. If the mice failed to press the reinforced lever within 60 s after the lever extension, both levers were retracted and the ITI started without delivering the reward; this was thus considered an incomplete trial. Representative behavior patterns for the FR1 trial are shown in ***Fig. 1B***, left side. A single session continued until 100 trials or one hour, which was conducted once a day for seven consecutive days (***Fig. 1C***, left side). A FR1-First session was defined as the session in which the mice first accomplished more than 50 completed trials. Establishment of the response–outcome relationship during the seven FR1 sessions was confirmed by an increase in the number of completed trials (***Fig. 1D***, left; FR1-First vs. FR1-Last, Tukey-Kramer, $p < 10^{-5}$). Development of the mice's FR1 task engagement was evident by a decrease in 1) the mice's reaction time between the lever extension to the first lever press (first press latency; ***Fig. 1D***, center;

FR1-First vs. FR1-Last, Tukey-Kramer, $p < 10^{-5}$) and 2) the number of incomplete trials per session (***Fig. 1D***, right; FR1-First vs. FR1-Last, Tukey-Kramer, $p < 0.01$).

After the establishment of the lever press–reward delivery relationship by the seven FR1 sessions, we increased the number of lever presses required for the food reward delivery from one to five. This schedule transition from FR1 to FR5 temporarily decreased the number of completed trials (***Fig. 1D***, left; FR1-Last vs. FR5-First, Tukey-Kramer, $p < 10^{-5}$), but the number was again significantly increased at the end of the FR5 schedule (***Fig. 1D***, left; FR5-First vs. FR5-Last, Tukey-Kramer, $p < 0.05$), implying behavioral adjustment in the mice during the six consecutive FR5 sessions. The behavioral measures for task engagement (first lever press latency and number of incomplete trials) and exploratory behavior (number of non-reinforced lever presses) did not change with the schedule transition (***Fig. 1D***, center and right; Tukey-Kramer, first press latency, $p = 0.62$; number of incomplete trials, $p = 0.77$; ***Fig. S1A***) nor throughout the FR5 sessions (Tukey-Kramer, first press latency, $p > 0.99$; number of incomplete trials, $p = 0.34$). These results prompted us to use another behavioral measure to detect the mice's behavior adjustment during the FR5 schedule.

To quantify the session-by-session adaptation of the mice to the FR5 schedule, we examined the completion latency, defined as the duration from the first lever press to the last (fifth) lever press in the completed trials. A gradual decrease in the completion latency with each FR5 session was observed, indicating that the mice performed the lever-press task more smoothly during the FR5 schedule (***Fig. 1E***; one-way repeated measures ANOVA, $p < 10^{-6}$). The percentage of incomplete trials per session were negligible after the mice had established the lever press–reward delivery relationship during the FR1 schedule ($1.5 \pm 0.8\%$ and $3.5 \pm 1.5\%$ of all trials in the FR1-Last and FR5-First sessions, respectively, $n = 8$ mice). Taken together, these findings showed that 1) once establishing an action–outcome relationship, the mice maintained their degree of engagement even after the ratio requirement was changed from FR1 to FR5 (***Fig. 1D***, center and right), and 2) the mice adjusted themselves to the FR5 sessions to efficiently attain beneficial outcomes by shortening their task completion duration (***Fig. 1E***).

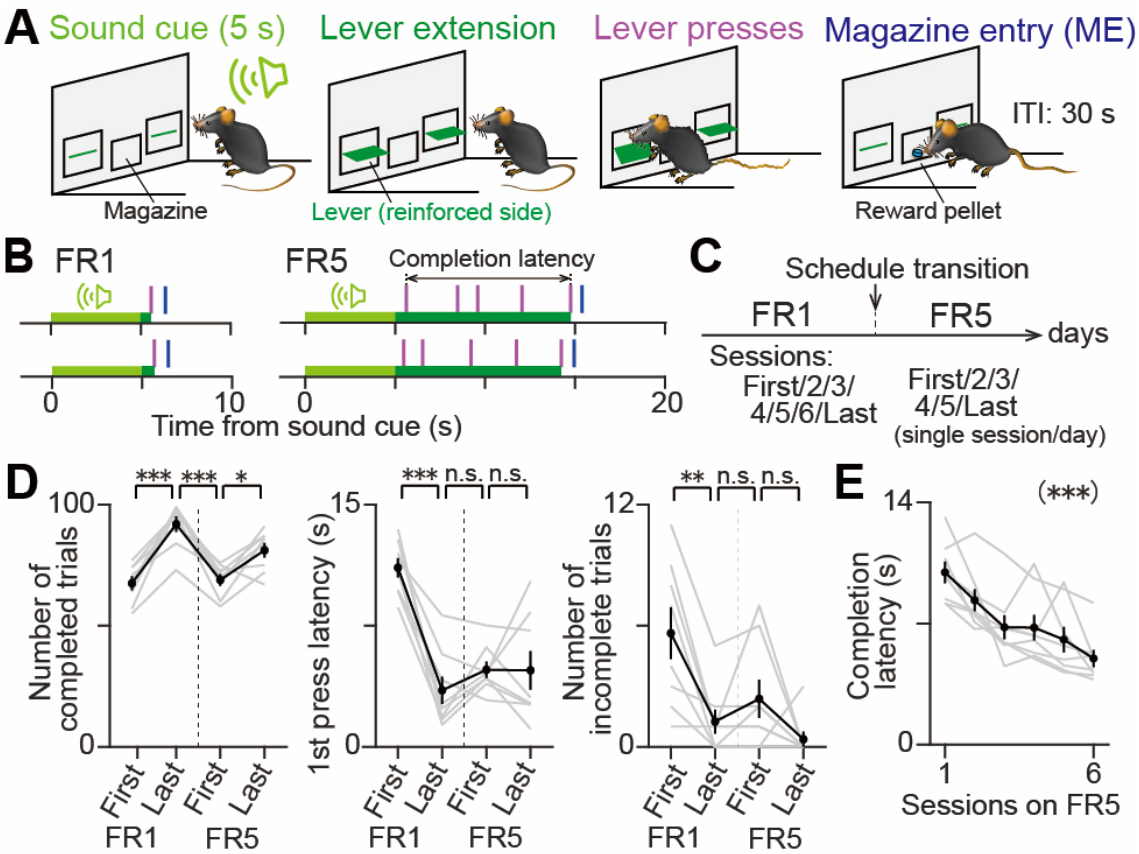

Figure 1. Behavior adjustment across FR5 sessions manifested in completion latency.

(A) Procedure of a fixed ratio lever press task. Only the left lever was reinforced; pressing the left lever for a predetermined number of times triggered the delivery of a food reward. If 60 seconds had elapsed from the lever extension onset without the pressing of the reinforced lever enough times, the levers were retracted and the ITI started without the delivery of a reward.

(B) Representative behavior patterns of individual trials on FR1 (left column) or FR5 (right column) schedules that required pressing the left lever once or five times for a reward, respectively. In each panel, the time axis was aligned with the sound cue onset. Light green bar, sound cue period (0–5 s). Dark green bar, lever extension period (starting 5 s after sound cue onset until the pressing of the last lever required for reward delivery). Vertical magenta tick, lever press. Vertical blue tick, magazine entry (ME).

(C) Mice were first trained on the FR1 schedule. Seven consecutive FR1 sessions, each including more than 50 completed trials, were designated as FR1-First, FR1-2, FR1-3, etc. After the FR1 sessions, six FR5 sessions followed. Vertical dashed line, schedule transition from FR1 to FR5.

(D) Session-by-session comparison: number of completed trials (left), first lever press latency (center), and number of uncompleted trials per session (right). It is evident that the association between the stimulus (the lever press) and outcome (the reward) becomes stronger during the FR1 sessions (n = 4 sessions with 8 mice each, Tukey-Kramer post-hoc analysis, FR1-First versus FR1-Last, $p = 4.2 \times 10^{-6}$ [number of completed trials], $p = 2.0 \times 10^{-6}$ [first press latency], $p = 6.2 \times 10^{-3}$ [number of incomplete trials]). While the

number of completed trials transiently decreased with a schedule transition from FR1 to FR5 (vertical dotted lines in ***Fig. 1D***), the first press latency and the number of incomplete trials failed to detect behavior adjustment during the FR5 sessions (FR1-Last versus FR5-First, $p = 9.6 \times 10^{-6}$ [number of completed trials], $p = 0.62$ [first press latency], $p = 0.77$ [number of incomplete trials]; FR5-First versus FR5-last, $p = 1.1 \times 10^{-2}$ [number of completed trials], $p > 0.99$ [first press latency], $p = 0.34$ [number of incomplete trials]). * $p < 0.05$. ** $p < 10^{-2}$. *** $p < 10^{-5}$.

(E) Completion latency, the time from the first to last lever press (***Fig. 1B***, right), detected behavior adjustment during FR5 sessions (n = 6 sessions with 8 mice each, one-way repeated measures ANOVA, $F_{(5,35)} = 13.18$, $p = 3.0 \times 10^{-7}$). Gray line, individual mouse. Black line, mean ± SEM. *** $p < 10^{-6}$.

Results 2: Magazine-Checking Behavior Quantifies Increasing Expectations for Task Completion

We next investigated the behavioral factors that accounted for the gradual decrease in the completion latency. We noticed that the mice occasionally checked the food magazine without the reward before completing the FR5 task during an inter-press interval between two consecutive lever presses (***Fig. 2A and 2B***). Because 99.0% of the checking behavior consisted of a single head entry to the magazine (4046 out of 4088 total cases from 8 mice), we defined the first head entry during an inter-press interval as a premature magazine entry (PME; red arrows in ***Fig. 2C***). After a PME, mice returned to the lever zone for a next press. We categorized the inter-press intervals into four groups according to the lever press order (e.g., $\text{Interval}_{2\text{–}3}$ represents the inter-press interval between the second and third lever presses) and examined the effect of the presence of a PME on the length of the inter-press interval, which constitutes the completion latency.

We observed that the inter-press intervals with PMEs were significantly longer than those without PMEs in each inter-press interval group (***Fig. 2D***; Kolmogorov-Smirnov test, all $p < 10^{-10}$), implying that a PME delays a task completion. Indeed, most variations in the inter-press intervals were explained by the presence or absence of a PME rather than by the session number or inter-press interval group (***Fig. 2E***). Note that the number of head entries during the ITI or reward periods did not change across the FR5 sessions (***Fig. S2-1A, B, C***). Therefore, the decrease in the number of PMEs was critical for the reduction in the completion latency across the FR5 sessions (***Fig. 1E***). We evaluated a behavior adaptation of a mouse to the FR5 schedule by examining PMEs in each inter-press interval group (***Fig. 2F***). The number of trials with a PME increased along with the inter-press interval group, implying that the mice became more expectant of a food reward during later inter-press intervals.

This interpretation assumes that the mice associated multiple lever presses with a reward in a FR5 schedule, instead of just associating a lever pressing with a reward. To verify this idea, we compared the number of lever presses before the first PME, with a session

number of a FR5 schedule (***Fig. S2-1D***). The average number of lever presses before the first PME increased gradually with the FR5 sessions, demonstrating that the mice learned to press a lever multiple times for a reward. We also compared the number of lever presses before the first PME using the real and shuffled data, in which the interval group number of a PME was shuffled (e.g., a PME of $\text{Interval}_{1–2}$ was randomly assigned to a PME of $\text{Interval}_{3–4}$). The number of lever presses before the first PME using the real data was significantly larger than that of the shuffled (1000 times) data in almost all sessions of all mice (n = 8, ***Fig. S2-1E, F***). These results support the idea that mice did not perform a PME randomly but based on their recognition of the lever press number with expectation for a reward.

The percentage of an inter-press interval with a PME in all the completed trials (PME percentage) was a sensitive measure for a behavior adjustment during the FR5 sessions. The PME percentage decreased with each session (***Fig. 2G***, early and late sessions indicated by black and brown lines, respectively) regardless of the inter-press interval group (***Fig. 2G***, x-axis) while maintaining the tendency of a larger PME percentage with later inter-press interval groups such as $\text{Interval}_{3–4}$ or $\text{Interval}_{4–5}$ (***Fig. 2G***, rising lines; two-way ANOVA, inter-press interval group, $p < 10^{-4}$; session, $p < 10^{-6}$). We quantified whether a PME occurred more frequently in the first half ($\text{Interval}_{1–2}$ and $\text{Interval}_{2–3}$) or second half ($\text{Interval}_{3–4}$ and $\text{Interval}_{4–5}$) of the interval groups during the FR5 sessions by calculating a PME bias; The PME bias was closer to –1 or 1 if a PME occurred more frequently in the first or last half of the interval groups, respectively; The PME bias was 0 if a PME occurred evenly in the first and last half of the interval groups. Plotting the mean PME percentage across all inter-press interval groups in a single FR5 session (Overall PME (%)) vs. the PME bias showed that the PME bias was near 0 in the early sessions and gradually shifted toward 1 in the later sessions (from black to brown open circles in ***Fig. 2H***; one-way repeated measures ANOVA, $p < 10^{-7}$), confirming that the overall PME percentage decreased gradually with the FR5 sessions (***Fig. 2H***, y-axis; one-way repeated measures ANOVA, $p < 10^{-7}$).

We selected a single session from each animal that exhibited a major behavior adjustment, which we used for subsequent analysis in this paper (hereafter, these sessions will be referred to as the selected FR5 session). The session selection was based on three criteria: 1) the number of PMEs in each inter-press interval group was 7 or more (***Fig. S2-2A***), 2) the number of the interval groups that accompanied decrease in the number of PMEs between the initial and last 20 trials was the largest among the six FR5 sessions. Typically, the number of the interval groups were three to four in a FR5 session (***Fig. S2-2B***), and 3) if there were multiple sessions that satisfied the previous two criteria, the session with the largest reduction rate in the number of PMEs between the initial and last 20 trials was chosen. The selected FR5 sessions reproduced our previous results (***Fig. S2-2A vs Fig. 2F; Fig. S2-2C vs Fig. 2H***). Mice stayed at the food magazine longer after a ME than

a PME, reflecting a consumptive behavior of a mice at the magazine after a ME (***Fig. S2-2D***).

Because PMEs constitute a reward-seeking behavior, we estimated a behavior-based reward expectation of a mouse by converting the PME percentage in each inter-press interval group of the selected sessions to a ratio. We compared the relationship between the reward expectation and inter-press interval groups, finding that the behavior based-reward expectations increased with the inter-press interval groups (***Fig. 2I***; one-way repeated measures ANOVA, $p < 10^{-10}$). This result corroborates our interpretation that reward expectation of a mouse increased nearer the task completion.

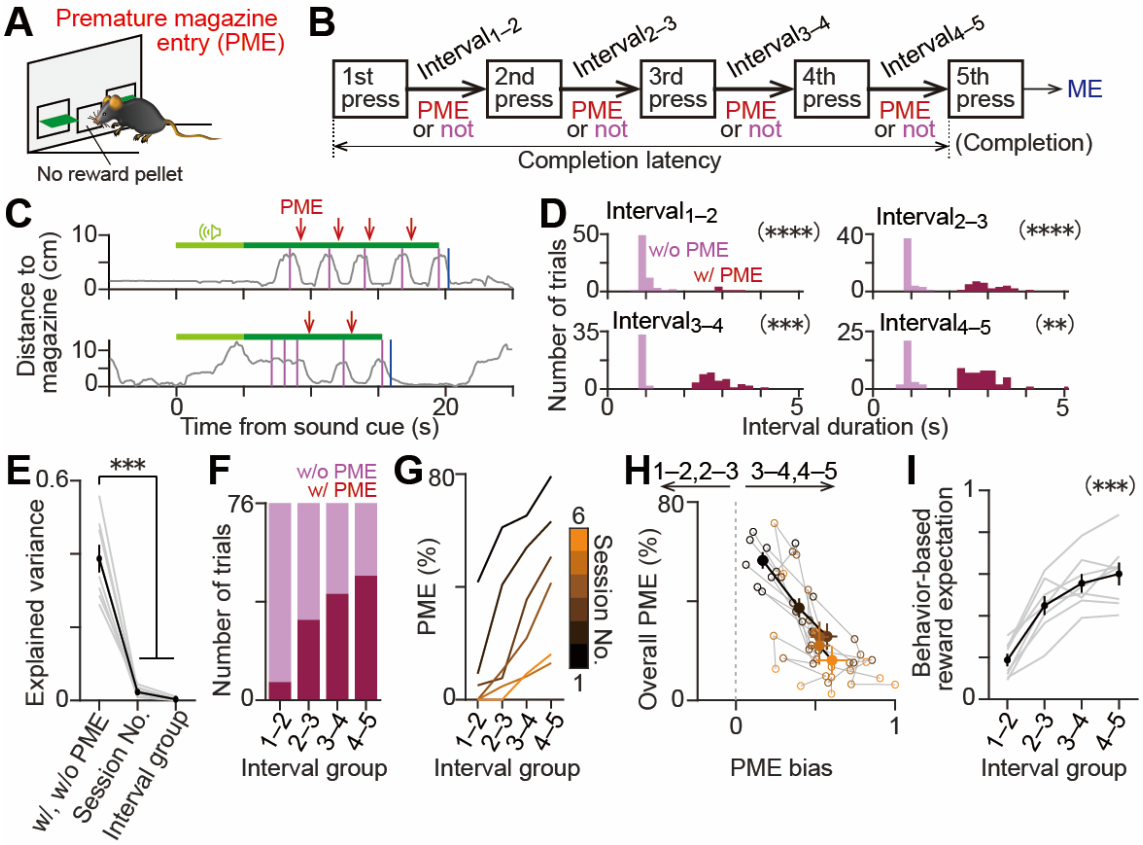

Figure 2. Rising expectations nearer task completion can be quantified by the percentage of a premature magazine entry.

(A) The mice occasionally checked the food magazine without a reward before completing the FR5 task, which we called premature magazine entry (PME).

(B) Scheme of a FR5 lever press task. An inter-press interval, a period between two consecutive lever presses, was distinguished by the order of a lever press, e.g., $\text{Interval}_{1\text{–}2}$ is the period between the first and second lever presses.

(C) Representative behavior patterns on a FR5 schedule with PMEs (red arrows) from mouse 9. A PME was detected based on the distance from the magazine to the mouse's head. Light green bar, sound cue period. Dark green bar, lever extension period. Vertical magenta tick, lever press. Vertical blue tick, ME.

(D) Representative histograms of inter-press interval durations, which were grouped into $Interval_{1–2}$, $Interval_{2–3}$, $Interval_{3–4}$, or $Interval_{4–5}$, with (dark red) or without (pale red) PME in a session of mouse 9. Histogram of inter-press interval durations with a PME was significantly longer than that without a PME ($Interval_{1–2}$, n = 30 and 42 events with and without PME, respectively, Kolmogorov-Smirnov test, $p = 6.1 \times 10^{-16}$; $Interval_{2–3}$, n = 44 and 28 events, $p = 2.3 \times 10^{-16}$; $Interval_{3–4}$, n = 47 and 25 events, $p = 1.2 \times 10^{-15}$; $Interval_{4–5}$, n = 57 and 15 events, $p = 1.1 \times 10^{-11}$). ** $p < 10^{-10}$. *** $p < 10^{-14}$. **** $p < 10^{-15}$.

(E) Explained variance analysis revealed that the variability of the inter-press interval durations was explained more by the presence or absence of a PME than by the session number (First/2/3/4/5/Last sessions) or inter-press interval group number ($Interval_{1–2/2–3/3–4/4–5}$) (n = 3 factors with 8 mice each, Tukey-Kramer post-hoc analysis, PME versus session number, $p = 4.3 \times 10^{-8}$; PME versus inter-press interval group number, $p = 2.3 \times 10^{-8}$). Gray line, individual mouse. Black line, mean ± SEM. *** $p < 10^{-7}$.

(F) A representative stacked bar chart of trials with (dark red) or without (pale red) a PME during the FR5-2 session of mouse 9. The mouse exhibited more PMEs as the FR5 task completion neared (e.g., the number of trials with a PME during $Interval_{4–5}$ was larger than that during $Interval_{1–2}$).

(G) Representative session-by-session plot of the PME percentage versus the inter-press interval groups during the FR5 schedule of mouse 9. While the mean of the PME percentage decreased as the session number increased (early and late sessions indicated as black and brown lines, respectively), the PME percentage within a session increased as the interval number increased, e.g., the PME percentage during $Interval_{4–5}$ was larger than that during $Interval_{1–2}$ (n = 24 data points, 4 interval groups × 6 sessions, two-way factorial ANOVA, interval number, $F_{(3,23)} = 18.42$, $p = 2.7 \times 10^{-5}$, session number, $F_{(5,23)} = 26.13$, $p = 6.7 \times 10^{-7}$).

(H) Group data plot shows an increase in a PME bias from the first (black) to the last (brown) sessions (n = 6 sessions with 8 mice each, one-way repeated measures ANOVA, $F_{(5,35)} = 16.63$, $p = 2.1 \times 10^{-8}$) and a decrease in the overall PME percentage ($F_{(5,35)} = 17.21$, $p = 1.4 \times 10^{-8}$) during the FR5 schedule. Open circle, individual session (session numbers are color-coded as in G). Filled circle, mean ± SEM within a session. Gray line, individual mouse. Black line, mean of all mice.

(I) Behavior-based reward expectations of mice increased as the FR5 task neared completion (n = 4 interval groups with a single session each from 8 mice, one-way repeated measures ANOVA, $F_{(3,21)} = 71.25$, $p = 3.6 \times 10^{-11}$). Gray line, individual session. Black line, mean ± SEM. *** $p < 10^{-10}$.

Results 3: DA Response upon ME is Modulated by the Schedule Transition from FR1 to FR5

To clarify the DA dynamics during the behavior adjustment, we recorded the extracellular DA fluctuation. We used a viral approach (PHP.eB AAV-hSyn-GRAB-DA2m-W, 1.0×10^{14} genome copies/ml) to express a genetically encoded optical dopamine sensor—$GRAB_{DA2m}$—in the ventrolateral striatum (VLS) (***Fig. 3A***). We targeted the VLS because our previous studies demonstrated that the VLS played a critical role in lever pressing operant tasks under FR schedules (***Natsubori et al., 2017; Tsutsui-Kimura et al., 2017a; Tsutsui-Kimura et al., 2017b; Yoshida et al., 2020***). We developed a fiber photometry

system to continuously monitor extracellular dopamine level fluctuations *in vivo* through an optical fiber during the operant task (***Fig. 3B***). The $GRAB_{DA2m}$ sensor was alternately excited by 405 nm and 465 nm lasers at 40 Hz, and the emissions were detected by a GaAsP photocathode with a time resolution of 50 ms. The fluorescence intensity ratio (465 nm/405 nm) captured the extracellular DA fluctuations on a timescale of sub-seconds (***Fig. 3C***). The fluorescence intensity was stable for an hour with little photobleaching effects, enabling consistent DA measurements for a session (***Fig. S3B, C***). Our fiber photometry system provided a direct measurement of extracellular DA fluctuations at a target site of DA neurons, complementing previous studies that investigated the activity of midbrain DA neurons with electrophysiological, electrochemical, or optical measurements (***Amo et al., 2020; Schultz et al., 1997; Starkweather and Uchida, 2021***). It has been reported that the expression of GRAB sensors did not change its downstream intracellular signaling (***Sun et al., 2018***). We confirmed that neither $GRAB_{DA2m}$ expression nor the optic fiber implantation in our study altered mice's behavior (***Fig. S1B, C; Fig. S2-1B, C***).

We succeeded in observing phasic increases in the extracellular DA levels at the onsets of the sound cue, lever extension, and ME in the FR operant task (***Fig. 3D and Fig. 3–video 1***). These DA dynamics were almost in line with previous studies in three aspects of Pavlovian and operant conditioning (***Amo et al., 2020***) (***Schultz et al., 1997***): First, it is reported that there is a large DA response to an unconditioned stimulus (US) and an almost negligible response to a conditioned stimulus (CS) at the beginning of the CS–US association. In our study, there were a clear DA response upon the US (a gray trace for DA dynamics upon a ME during FR1-First in the bottom right panel of ***Fig. 3E***) and relatively small DA responses upon the CSs (gray traces for DA fluctuation upon the sound cue and lever extension during FR1-First in the bottom left panel of ***Fig. 3E***) during the FR1-First session. Secondly, it is reported that a DA activity shows clear responses to the CS but no obvious responses to the US once the CS–US association is established. In our study, there were a significant increase of DA responses upon the CSs (a black trace for DA responses upon a sound-cue [green bar] or lever extension [dark green bar] in the bottom left panel in ***Fig. 3E***; group data comparison of DA response amplitudes between FR1-First vs FR1-Last upon a sound cue in ***Fig. 3F*** and a lever extension in ***Fig. 3G***; Tukey-Kramer, $p < 10^{-3}$), and a significant decrease of a DA response upon the US (a black trace for DA fluctuation upon a ME at bottom right panel in ***Fig. 3E***; group data comparison of DA response amplitudes upon a ME between FR1-First vs FR1-Last in ***Fig. 3H***; Tukey-Kramer, $p < 0.05$) from the FR1-First to FR1-Last sessions. And thirdly, it is reported that DA responses to the US and CS increases and diminishes, respectively, returning to the initial state of the CS–US association upon a violation of the CS–US association. In our study, there were an augmentation of a DA response upon the US (black versus light purple traces upon a ME in the bottom right panel of ***Fig. 3E***; group data comparison of DA response amplitudes upon a ME between

FR1-Last vs FR5-First in ***Fig. 3H***, Tukey-Kramer, $p = 2.3 \times 10^{-7}$), and a decrease of a DA fluctuation upon the CS (black versus light purple traces for DA fluctuation upon a lever extension [dark green bar] at left panels in ***Fig. 3E***; group data comparison of DA response amplitudes upon a lever extension between FR1-Last vs FR5-First in ***Fig. 3G***, Tukey-Kramer, $p < 0.05$) from the FR1-Last to FR5-First sessions.

The above results demonstrate that 1) an association between a food reward and lever pressing was established during the FR1 sessions, and 2) the schedule transition from the FR1-Last to FR5-First sessions violated the CS–US association. A minor difference of DA dynamics between our observation and the previous studies was that the DA response to another CS (sound cue) did not diminish and remained large even after the schedule transition (***Fig. 3F***, FR1-Last versus FR5-First; Tukey-Kramer, $p = 0.96$). This observation suggests that the schedule transition did not diminish completely the association between the CS–US association, which might explain why the mice maintained their motivation and continued to perform operant responses even after the schedule transition (***Fig. 1D***, 1st press latency and number of incomplete trials between FR1-Last vs FR5-First). Note that DA increase upon lever pressing was not significant throughout our study (magenta lines in ***Fig. 3D***), which was consistent with a previous report that activity of DA axons projecting to the ventral striatum was not as modulated by mice's motion as that projecting to the dorsal striatum (***Howe and Dombeck, 2016***).

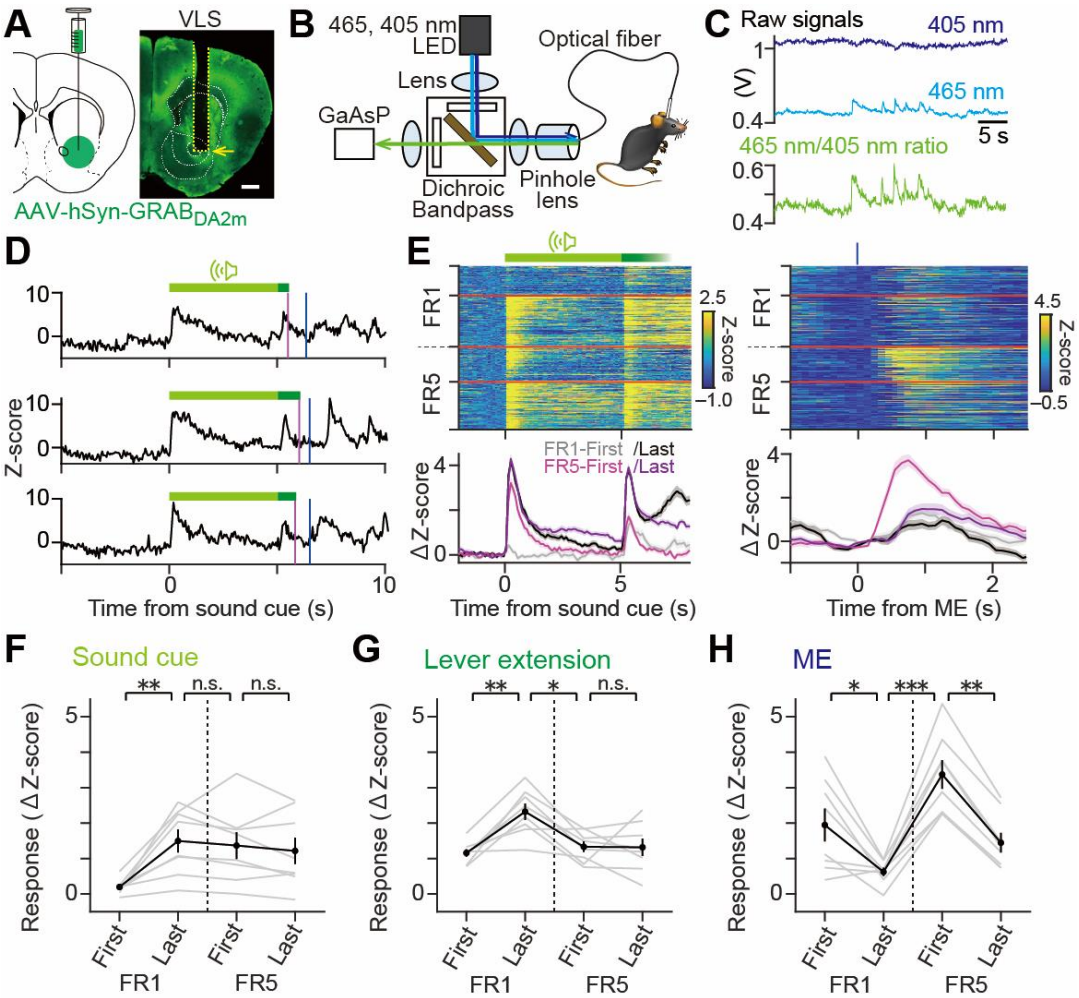

Figure 3. Optical measurement of extracellular DA dynamics in the ventrolateral striatum during a fixed ratio lever press task.

(A) Left: Schematic illustration of AAV virus injection in the VLS for the expression of a fluorescent DA sensor, GRAB$_{DA2m}$. Right: Expression pattern of GRAB$_{DA2m}$ (green) in an example mouse and the location of the optic fiber in the VLS. Yellow dashed line, optical fiber track. Yellow arrow, fiber tip. Scale bar, 500 μm.

(B) Schematic of the fiber photometry system. The light path for fluorescence excitation and emission is through a single multimode fiber connected to the optical fiber implanted in the VLS. Excitation light was applied continuously with alternating wavelengths of 405 nm (dark blue arrow) and 465 nm (blue arrow) at 40 Hz. Fluorescence emissions at 525 nm (green arrow) were detected using a GaAsP photomultiplier tube.

(C) Example fluorescence intensity fluctuations of GRAB$_{DA2m}$ excited by 465 nm (blue line) and 405 nm (dark blue line) light. Corresponding ratio values (green; 465 nm/405 nm) indicate extracellular DA dynamics.

(D) Representative GRAB$_{DA2m}$ signal dynamics during the last session of the FR1 schedule (FR1-Last) of mouse 6. A DA surge was observed after the onset of the sound cue (light green bar) and lever extension (dark green bar). A phasic DA increase also followed an ME (vertical blue line) after a reinforced lever press (vertical magenta line).

(E) Top: Representative heatmaps demonstrating GRAB$_{DA2m}$ signal fluctuations during sound cue periods (light green bar in the left panel) followed by lever extension onsets (dark

green bar in the left panel) or ME timings (vertical blue line at the top of the right panel) in four selected sessions (from top to bottom, FR1-First, FR1-Last, FR5-First, and FR5-Last sessions). Signals were centered to that from –1 to 0 s at the sound cue onset. The horizontal gray dashed line at the left axis indicates the schedule transition from FR1 to FR5. Bottom: Averaged $GRAB_{DA2m}$ signal traces for respective sessions on the heatmap.

(F—H) $GRAB_{DA2m}$ signal dynamics at lever extension (G) or ME (H) captured the schedule transition from FR1 to FR5 (n = 4 sessions with 8 mice each, Tukey-Kramer post-hoc analysis: FR1-Last versus FR-First, $p = 2.6 \times 10^{-3}$ [lever extension], $p = 2.3 \times 10^{-7}$ [ME]), while that at the sound cue onset (F) did not (n = 4 sessions with 8 mice each, Tukey-Kramer post-hoc analysis: FR1-Last versus FR5-First, p = 0.96 [sound cue]). Gray line, individual mouse. Black line, mean ± SEM. * $p < 0.01$. ** $p < 10^{-3}$. *** $p < 10^{-6}$.

Results 4: DA Dip Amplitudes Quantify Disappointment

We analyzed the DA dynamics during the PME in detail because the PME ratio reflected the reward expectations of a mouse (***Fig. 2I***) and PMEs were the only cue for mice to realize a schedule transition from FR1 to FR5. For later analysis, PMEs were classified into four groups according to the lever press order, e.g., a $PME_{2–3}$ is a PME that occurred between the second and the third lever presses (***Fig. 4A***). We noticed a transient drop in the DA levels (***Fig. 4B and Fig. 4–video 1***, black arrows just above the Z-scored DA fluctuation) following a PME (***Fig. 4B***, red arrows at the top) of a representative mouse. The mean time-course of the DA fluctuation of all trials in the selected sessions of the mouse demonstrated the phasic decrease of a DA response upon a PME (DA dip; ***Fig. 4C*** bottom, red trace; PME_{all} includes all the four PME groups), in addition to the phasic DA increase upon a ME (DA surge; ***Fig. 4C*** bottom, blue trace). It is noteworthy that the same head entry behavior during the ITI periods, which were the lowest possible state of reward expectation in the FR5 task, did not evoke DA fluctuation (***Fig. 4C***, black trace). Furthermore, the number of head entries during ITIs did not change across FR5 sessions (***Fig. S2-1B, C***, ITI). Statistical analysis confirmed the above DA dynamics (***Fig. 4D*** for the representative mouse; ***Fig. 4E*** for all mice). These results indicate that DA fluctuation upon a head entry depended on the animals' internal state reflecting degree of reward expectation.

We found that amplitudes of a DA dip within a session increased with the PME groups (***Fig. 4F*** and ***Fig. S4-2*** for DA dip traces of the representative mouse and all the mice, respectively; ***Fig. 4G*** for a group data comparison of DA dip amplitudes), implying that proximity to the completion of the task affected the amplitude of the DA dip. In these analyses, a DA dip was calculated as a difference of z-scores from the magazine entry (Δ z-score), assuming a constant pre-dip DA level and emphasizing the role of phasic DA activities (***Fig. 4I***; a dip-trough scenario). Another possibility for realizing the variation of the DA dip amplitudes was a modulation of a tonic DA level preceding the DA dip with a constant DA dip-trough level (***Fig. 4H***; a pre-dip scenario), which emphasized the role of tonic DA fluctuations for achieving a quantitative representation of a negative RPE (***Hamid et al., 2016***). We assessed the two scenarios by analyzing DA level variations at

pre-dip and dip-trough periods (***Fig. 4H, I***); Z-scores of the DA signal was calculated using the ITI periods as a baseline because DA signals during the ITI periods were stable among the PME groups (***Fig. S4-5C***). We revealed that both scenarios were valid in our study. DA levels both at pre-dip and dip-trough periods showed increasing and decreasing trends, respectively, with PME groups (***Fig. 4J, K***). These results support the larger DA dip amplitudes with the mice's increasing proximity to task completion.

We discovered a significant positive correlation between an amplitude of a DA dip upon a PME and the magnitude of a behavior-based reward expectation (***Fig. 4L***). We hypothesized that the DA dip size upon a PME quantified the magnitude of disappointment because 1) disappointment is considered as the realization that an actual outcome felled below expectations (***Steiner and Redish, 2014***), 2) the RL theory formulates the RPE as the difference in the achieved reward and the expected reward, 3) DA dynamics reflect the RPE, and 4) the achieved reward after the PME was zero (no reward) in the current study. It is noteworthy that a DA surge upon a FR5 task completion was unlikely to be involved in behavioral refinement by reducing a PME behavior because an amplitude of DA surges upon ME did not distinguish the number of a PME in a task (***Fig. S4-4F***), indicating that behavior refinement in the FR5 task depended on a disappointment upon PMEs rather than successes upon MEs.

A DA dip upon a PME was consistent and reproducible. Z-scored DA signal 0.5−1.5 s after a PME was below zero in 90.3 ± 2.0% of all PMEs (***Fig. S4-1E, F***), and the DA signal decreased upon a PME in 89.9 ± 1.4% of all PMEs (***Fig. S4-1G, H***). The time interval 0.5−1.5 s after a PME was selected because the negative peak of the DA dip occurred on average 1.09 ± 0.03 s (n = 8 mice) after a PME (***Fig. S4-1C, D***). This delay includes approaching behavior to the magazine because a PME was detected when mice came within 2.5 cm of the magazine. Therefore, we analyzed a locomotive speed of mice upon a PME and found a stereotypical behavior of mice that arrived at and started to explore the magazine 0.37 ± 0.02 s after the PME detection (***Fig. S4-1A, B***). The frequency per session of PMEs decreased with FR5 sessions (***Fig. S2-1B, C***, IPI; ***Fig. S2-1G***), just as the mean amplitudes of a DA surge upon the MEs (***Fig. S4-1I***, right), corresponding to the establishment of the CS–US association (***Fig. 1E***). The mean amplitudes of a DA dip upon the PMEs, by contrast, did not change with the sessions (***Fig. S4-1I***, left), implying that DA dip amplitudes might reflect mice's recognition that the expected reward was not obtained. The frequencies per minute of PMEs (paired with a DA dip) and MEs (paired with a DA surge) were 1.9 ± 0.1 and 1.2 ± 0.03, respectively, in the selected FR5 sessions (***Fig. S4-1J, K, L***). The frequencies per trial of PMEs were 1.8 ± 0.2 in the selected FR5 sessions.

We considered two possibilities against our interpretation that an amplitude of a DA dip increased with PME groups. The first one assumes the two: 1) amplitudes of a DA dip

increased (or decreased) with trials within a session regardless of the PME groups, and 2) only PMEs of larger PME groups appeared more in later (or earlier) trials. This situation results in an apparent increase of DA dip amplitudes with PME groups even if there was no relation between a DA-dip depth and PME groups. To test this possibility, we compared the DA dip amplitudes in the first and the second half of a session, revealing no difference in DA dip amplitudes between them (***Fig. S4-3A, B***). In addition, the number of PMEs showed decreasing trends within a session regardless of the PME groups (***Fig. S4-3C, D***). These results did not support the first possibility. The second possibility is related to a waiting time from the mice's task engagement (the first lever press in a trial) to a PME. If there was a positive correlation between the waiting time and the amplitude of a DA dip, DA dip amplitudes would increase with PME groups even if DA dip amplitudes and PME groups were unrelated. To test this possibility, we calculated a correlation coefficient between the waiting time and DA dip amplitudes upon a PME (***Fig. S4-4A, B***). Note that the correlation coefficient was calculated for each PME group and that PME groups were further divided based on the number of preceding PMEs in the same trial, e.g., $PME_{3–4\,(0)}$ and $PME_{3–4\,(1)}$ were $PME_{3–4}$ without and with a preceding PME, respectively. These subdivisions were necessary to subtract the effects of the number of presses and PMEs on the waiting time because the waiting time naturally increases with the press number (***Fig. S2-2F***) and a PME (***Fig. S2-2G; Fig. 2D***). After these subdivisions, we found no correlation between the waiting time and DA dip amplitudes in more than 95% of the cases (***Fig. S4-4A and S4-4B***). These results did not support the second possibility. Therefore, we concluded that DA dip amplitudes increased with PME groups.

We examined other interpretations against our hypothesis that a DA dip size upon a PME quantified the magnitude of disappointment: regret and frustration. Regret is reported to occur in a situation where a given failure is due to one's own wrong action (***Dijk et al., 1999; Steiner and Redish, 2014; Zeelenberg et al., 2010***); If a PME was associated with regret rather than a disappointment, mice would reduce a PME behavior and increase lever-pressing. To examine if a PME led to a behavioral change in response to regret, we compared number of lever-presses before and after a PME (***Fig. S4-5A***). We found that the number of lever-pressing did not increase after a PME, a result against the regret interpretation. Frustration in rodents was quantified by duration of a lever press (***Vasquez et al., 2021***); If larger PME groups were associated with frustration, mice would increase duration from a lever pressing to an initiation of a PME. To test this possibility, we divided an inter-press interval into three periods: press−PME, PME−exit, and exit−press periods. We found a shorter press−PME period (***Fig. S2-2E***) and a longer stay at the magazine (***Fig. S4-4C, D***) with inter-press interval groups, results against the frustration interpretation. Taken together, our results demonstrate that DA dip amplitudes upon a PME reflect degree of disappointment in our FR5 schedule.

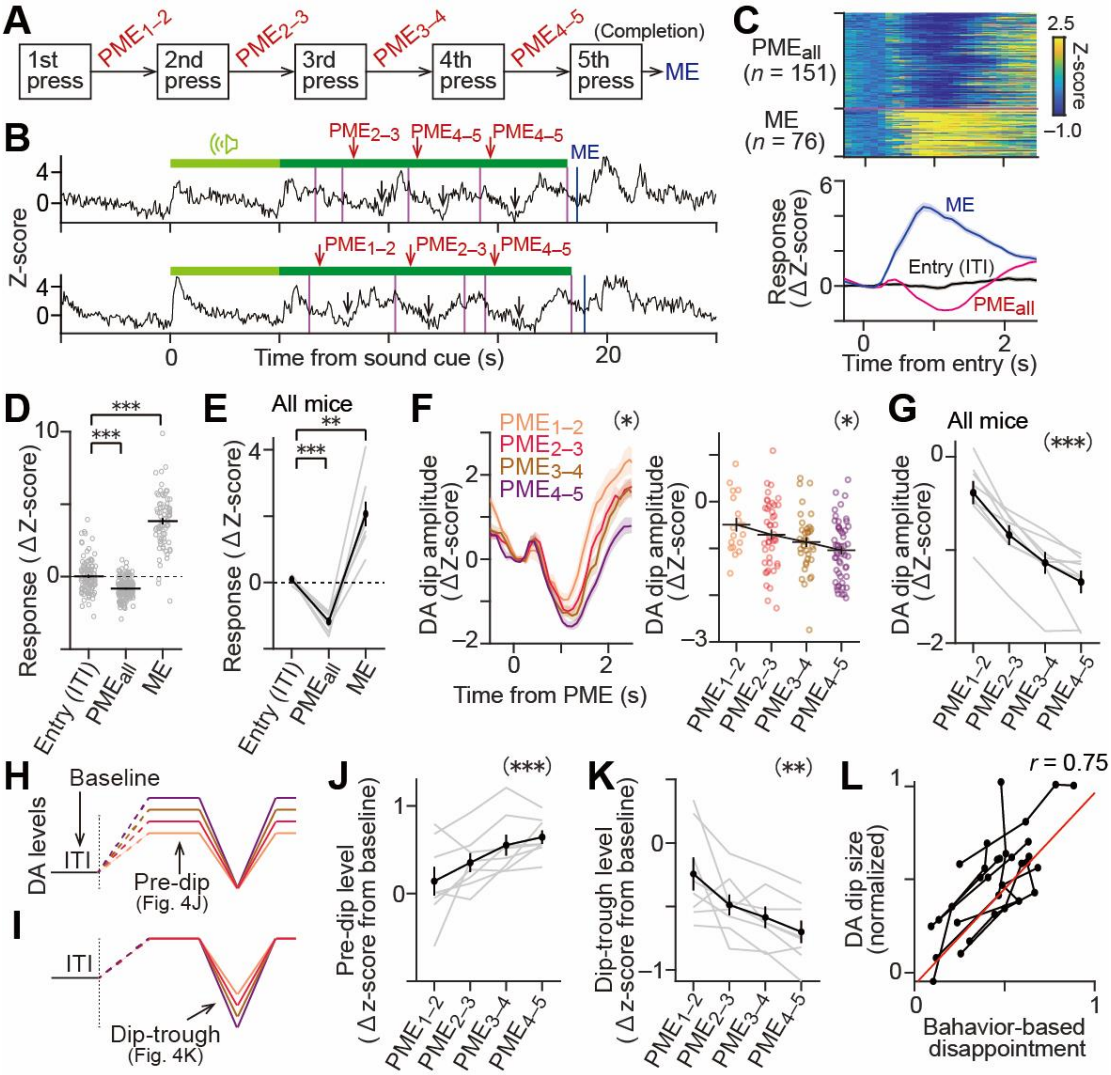

Figure 4. A DA dip amplitude reflects growing disappointment after a PME as completion of a FR5 task approaches.

(A) Premature entries (PMEs) were grouped by the order of lever presses, e.g., $PME_{1–2}$ was a PME that occurred during a period between the first and the second lever presses. ME, magazine entry.

(B) Two representative $GRAB_{DA2m}$ signal dynamics of mouse 11 during trials with PMEs on a FR5 schedule. A DA dip (black arrow) followed a PME (red arrow). Light and dark green bars, sound cue and lever extension periods. Magenta and blue lines, lever pressing and magazine entry (ME).

(C) Top: A representative heatmap of $GRAB_{DA2m}$ signal time courses aligned with the onset of the PMEs or MEs of mouse 11. PME_{all} includes every PME group such as $PME_{1–2}$ and $PME_{3–4}$. Bottom: $GRAB_{DA2m}$ signal traces (mean ± SEM) for PME_{all}, ME, and the head entry during ITI on the heatmap with a bin size of 100 ms. DA dip and surge following a PME (red) and ME (blue), respectively, were evident. A Δz-score indicates a difference of a z-score from the magazine entry by centralizing the original z-scores with their values at 0–0.2 s from the magazine entry.

(D) Comparison of a DA fluctuation for ITI, PME, or ME of mouse 11 in ***Fig. 4C***, calculated as a mean Δ z-score during 0.5–1.5 s after a magazine entry (Entry during ITI, n = 109 events, Δ z-score = 0.02 ± 0.11; PME_{all}, n = 151 events, Δ z-score = –0.83 ± 0.06; ME, n = 76 events, Δ z-score = 3.82 ± 0.21, Tukey-Kramer post-hoc analysis, entry during ITI versus base (zero), p > 0.99; entry during ITI versus PME_{all}, $p = 3.8 \times 10^{-9}$; entry during ITI versus ME, $p = 4.1 \times 10^{-9}$). Open circle, individual event. Black line, mean ± SEM. *** $p < 10^{-8}$.

(E) Group data of DA fluctuations confirmed a significant DA dip and surge after a PME and ME, respectively, when compared to a DA fluctuation upon a magazine entry during an ITI period (n = 8 mice each, Δ z-score = 0.08 ± 0.03 [entry during ITI], –1.17 ± 0.1 [PME_{all}], 2.07 ± 0.37 [ME], Tukey-Kramer post-hoc analysis, entry during ITI versus base (zero), p = 0.99; entry during ITI versus PME_{all}, $p = 2.0 \times 10^{-6}$; entry during ITI versus ME, $p = 8.0 \times 10^{-4}$). Gray line, individual mouse. Black line, mean ± SEM. ** $p < 10^{-3}$. *** $p < 10^{-5}$.

(F) Representative $GRAB_{DA2m}$ traces during the first session of a FR5 schedule (FR5-First) of mouse 11, plotted individually for each PME group. Left: An amplitude of a DA dip following a PME significantly increased when the PME group number increased, e.g., a DA dip following $PME_{4–5}$ was deeper than that following PME_{1-2}. Right: Same as the left panel but individual DA dip amplitude was plotted against PME groups (n = 18–52 events, one-way factorial ANOVA, $F_{(3,150)} = 3.87$, $p = 1.1 \times 10^{-2}$). * $p < 10^{-2}$.

(G) Group data comparison of DA dip amplitudes among PME groups confirmed deeper DA dips with increasing PME group number (n = 4 PME groups in a single session each from 8 mice, one-way repeated-measures ANOVA, $F_{(3,21)} = 39.41$, $p = 8.3 \times 10^{-9}$). Gray line, individual mouse. Black line, mean ± SEM. *** $p < 10^{-8}$.

(H, I) Both pre-dip (H) and dip-trough (I) scenarios can realize the larger DA dip size with PME groups. In the pre-dip scenario, modulation of a DA level during a pre-dip period (0–0.2 s from a PME) with a constant DA level during a dip-trough period (0.5–1.5 s after a PME) achieves the larger DA dip size with PME groups. In the dip-trough scenario, a DA level during a pre-dip period is constant but that during a dip-trough period varies to achieve the same outcome for the DA dip size. Line colors correspond to a PME group number as in ***Fig. 4F***. ITI period was used as a baseline of $GRAB_{DA2m}$ signal.

(J) DA levels during a pre-dip period was significantly different among PME group numbers, consistent with the pre-dip scenario (4 PME groups in a session per a mouse from 8 mice each, one-way repeated-measures ANOVA, $F_{(3,21)} = 8.22$, $p = 8.3 \times 10^{-4}$). Gray line, individual session. Black line, mean ± SEM. *** $p < 10^{-3}$.

(K) DA levels during a dip-trough period was significantly different among PME group numbers, consistent with the dip-trough scenario (one-way repeated-measures ANOVA, $F_{(3,21)} = 7.54$, $p = 1.3 \times 10^{-3}$). Gray line, individual session. Black line, mean ± SEM. ** p < 0.01.

(L) Variations of a DA dip size upon a PME (a difference in DA levels between pre-dip and dip-trough periods, ***Fig. 4F***) showed a positive correlation with the magnitude of disappointment (an inverse of a behavior-based reward expectation, ***Fig. 2I***; 4 PME groups in a session per a mouse from 8 mice each; Pearson correlation coefficient, $r = 0.75$, $p = 9.2 \times 10^{-7}$). Dots represent each data of a PME group, such as $PME_{3–4}$, in a session of a mouse. Lines connect data from the same session of a mouse.

Results 5: RL Model Reproduced Behavioral Adjustment and DA Dip Variation

We employed a computational modeling of a behavioral data to address whether the DA dip size variation had any functional role in behavior adjustment during the FR5 schedule. We used an actor-critic RL model with a temporal difference (TD) error because DA fluctuation has been proposed to represent a TD error, which was a teaching signal to guide experience-based behavior adjustment (***Amo et al., 2020; Moutoussis et al., 2009; O'Doherty et al., 2003; Schultz et al., 1997; Starkweather and Uchida, 2021; Suri and Schultz, 2001***). An agent of the model had a state value and action values for a forward step (lever-pressing) and an upward step (checking a magazine for a ME or PME), for each lever zone in a two-dimensional grid world environment (***Fig. 5A***). The agent began at the leftmost start zone in the environment and chose an action for a forward or upward step using a softmax action selection rule based on the action values. The agent updated the state and action values step-by-step using a state value-derived TD error. We assumed that the mice might occasionally misrecognize the current lever press number, e.g., mice might mistake the second lever press for the fourth one, because there was no external cue about the number. To introduce this stochastic error of the mice, we implemented a non-deterministic "state transition" (***Starkweather et al., 2018***) in our model; Even if the agent chose an action for a single forward step, the agent moved 1 to a maximum of 5 forward step(s) in the grid world (***Fig. 5B***, thick and thin black arrows). Similarly, the agent's choice of an action for a PME was followed by zero to a maximum of 5 backward step(s) (***Fig. 5B***, thick and thin orange arrows). We performed a grid search to fit eight free parameters of the model for the behavioral data of the mice during the FR5 schedule (***Fig. 5C***). The free parameters were the initial state value for a lever zone, the initial action values for a lever press and a PME, learning rates for state and action values, an inverse temperature for the softmax function, and forward and backward sigma values that specified the probabilistic state transition. Goodness of the fit was assessed by calculating a log likelihood (LLH) of action selection probabilities of a model for a lever press and a PME (***Fig. 5D,*** comparison of observed and fitted action selection probabilities).

We found that the best fitted models had a significantly higher LLH value than a random model that chose a lever press or a PME with an equal probability (***Fig. 5E***; one-sample t-test, LLH = -0.63 ± 0.01, $p < 10^{-9}$). The best fitted models also had significantly higher initial action value for a PME than that for a lever press (***Fig. 5F***; Wilcoxon signed rank test, $p < 0.01$), suggesting that the CS–US association during the FR1 schedule biased action values for a PME over a lever press during initial sessions of the FR5 schedule. The best-fitted models had sigma values larger than zero for the state transition, implying that mice actually misrecognized the current lever press number. We discovered that the absolute TD error of the best-fitted models increased with PME group numbers (***Fig. 5H***,

representative data; ***Fig. 5I***, group data), reproducing the larger DA dip amplitudes of the mice with PME group numbers (***Fig. 4F, G***). Moreover, the best-fitted models showed a positive correlation between the TD error size and the magnitude of disappointment that was estimated with PME percentages (***Fig. 5J***), corresponding to the positive correlation between a DA dip size and mice's behavior-based disappointment (***Fig. 4L***). These results of the computational modeling supported the functional relevance of a DA dip variation for the behavior adjustment during the FR5 schedule.

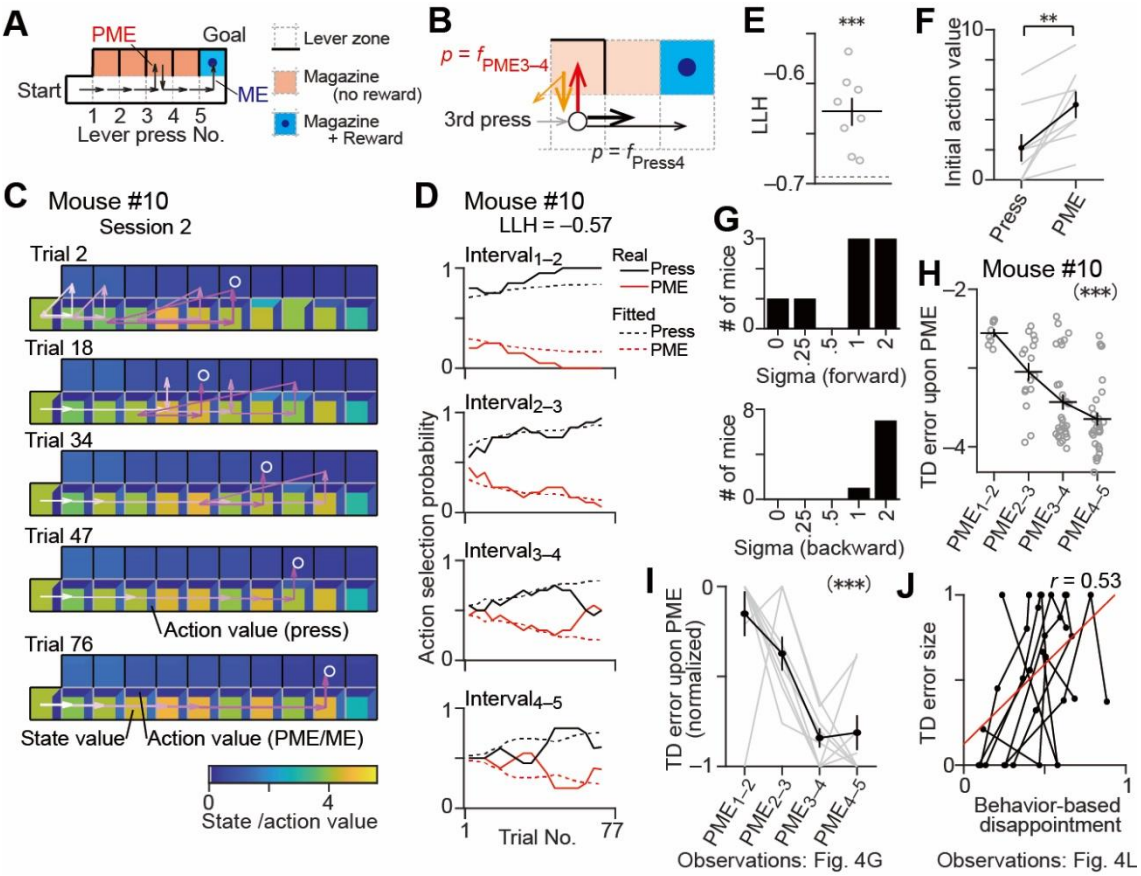

Figure 5. A computational model fitted to the mice's behavior replicates the increasing trend of DA dips near FR5 task completion.

(A) An actor-critic RL model with a TD error in a two-dimensional grid world environment was used to simulate behavioral data of mice in the FR5 schedule. An agent of the model began at the leftmost Start zone, moving step-by-step toward the top right Goal zone for a reward. Horizontal arrow, lever-pressing. Upward arrow, a magazine checking behavior for a PME (red) or ME (blue).

(B) A state transition was implemented in the model to simulate occasional misrecognition of the current lever press number by mice. In this example without a state transition, the agent of the model (open circle) chose either a $PME_{3–4}$ (red arrow) with a probability of $f_{PME3–4}$ returning to the same lever zone (thick orange arrow) or the fourth lever press (thick black arrow) with a probability of f_{press4} moving to the next lever zone. With a forward state transition, the agent mistook the fourth lever press with the fifth lever press skipping a lever zone (thin black arrow). With a backward state transition, the agent mistook the

third lever press number with the second lever press number (thin orange arrow) after the PME_{3-4}.

(C) Representative steps (arrows) of a model fitted to a behavioral data of a mouse #10. Order of the steps are indicated by graded colors from white (start) to magenta (goal). State values of the model are indicated as a color-coded square in each lever zone. Action values for a PME and a lever press of the model were color-coded at the top and right edges of the lever zone, respectively. Note that the position of the rewarded zone (white open circle) was variable due to the stochastic state transition, demonstrating that the agent did not necessarily follow the lever press number of actual mice behavior during the FR5 schedule.

(D) Comparison of time series of action selection probability during completed trials of a representative mouse (solid lines) and the best-fitted model (dashed lines). Action selection probabilities were averaged across 20 trials.

Group of plots for each inter-press interval group number comparing

(E) Group data of a LLH per action of the best-fitted models for each mouse. LLHs of the fitted models were significantly larger than that of a random model (dashed line) (n = 8 mice, one-sample t-test, LLH = -0.63 ± 0.01, $t_{(7)} = -47.47$, $p = 4.8 \times 10^{-10}$). *** $p < 10^{-9}$.

(F) The initial action value for PMEs was significantly higher than that for the lever presses in the best-fitted models (n = 8 mice, Wilcoxon signed rank test, signed rank = 0, $p = 7.8 \times 10^{-3}$), reflecting the preceding FR1 sessions. ** $p < 10^{-2}$.

(G) Group data of forward (upper panel) and backward (lower panel) sigma values for state transitions in the best-fitted models. Forward and/or backward sigma values were greater than zero, implying that mice occasional misrecognized the lever press number.

(H) Representative data of the best-fitted model for a behavioral data of Mouse #10, demonstrating that absolute TD error increased with PME groups (one-way ANOVA, $F_{(3,84)} = 14.66$, $p = 1.0 \times 10^{-7}$). Each open circle represents an individual PME event. Black line, mean ± SEM. *** $p < 10^{-6}$.

(I) Group data for normalized TD error amplitudes of the best fitted models confirm larger TD errors with PME group numbers (n = 4 PME groups in a single session each from 8 agents, one-way repeated-measures ANOVA, $F_{(3,31)} = 12.63$, $p = 2.1 \times 10^{-5}$), reproducing the results of mice (***Fig. 4G***). Gray lines, individual data. Black lines, mean ± SEM. *** $p < 10^{-4}$.

(J) The best-fitted models showed a positive correlation between the TD error size and the behaviorally estimated magnitude of disappointment, reproducing the positive correlation between a DA dip size and a magnitude of disappointment (***Fig. 4L***) (n = 4 PME groups in a single session each from 8 agents; Pearson correlation coefficient, $r = 0.48$, $p = 5.3 \times 10^{-3}$). Dots represent each data of the best-fitted model. Lines connect data from the same session of the models.

Discussion

We addressed the neuronal representation of disappointment. Our data from behavioral experiments, DA fluorescence observations, and computational modeling of behavioral data consistently supported the interpretation that the magnitude of the DA dip is a quantitative measure of disappointment. This interpretation is in line with previous reports that a DA level reduced transiently upon unreinforced head entries to the reward location without a reward during an operant task in a progressive ratio (PR) schedule (***Ko***

and Wanat, 2016), and that disappointment occurred in rodents when they encountered a reward omission (***Steiner and Redish, 2014***). Furthermore, analysis of our behavioral data favored the interpretation of disappointment upon a PME over other interpretations such as regret and frustration.

The simple operant task in this study had three advantages for our finding that DA dip amplitudes reflected magnitude of disappointment: 1) the fixed reward size, 2) the stereotypical behavior for a reward expectation, and 3) the modulation of a reward expectation in a single trial. The reward size was always a single pellet after the FR5 task completion. In other behavioral tasks, a reward size is commonly varied to modulate the magnitude of expectation, e.g., the amount of juice, which might introduce an extra DA fluctuation unrelated to disappointment. The PME in the FR5 task accompanied an almost identical behavior pattern for reward expectations of different magnitude. This situation contrasts with other tasks in which mice must choose distinct behaviors such as pressing left or right levers for a reward with different expectancy, which might cause extra DA fluctuations reflecting the behavior differences. The magnitude of reward expectations increased monotonically between the PME groups in a single trial, which is a unique task design when compared to other common tasks such as a lottery economic task, a random-ratio (RR) task, and a progressive ratio (PR) task. Although the latter tasks enable a flexible manipulation by an experimenter of reward expectation, multiple trials with or without rewards are necessary in these tasks to modulate the magnitude of reward expectations of mice, which might evoke a long timescale DA level fluctuation unrelated to disappointment at the time.

The combination of tonic and phasic fluctuation of extracellular DA level constituted the variation of the DA dip amplitude in our study. There have been two hypotheses for realizing the variations in a DA dip amplitude: tonic and phasic. The first hypothesis assumes that tonic DA modulation realizes variation in the DA dip amplitude (***Hamid et al., 2016***); A larger DA dip is apparently realized by a higher DA level prior to the dip while the DA concentration during the dip-trough is constant. Another hypothesis proposes that phasic DA fluctuation modulates DA dip magnitudes (***Dabney et al., 2020***); A larger DA dip is realized by variable DA level at the trough while the pre-dip DA level is constant. Our sensitive and stable DA imaging system combined with the simple FR5 operant task was critical to reveal that both tonic and phasic fluctuations of DA levels realized the variation of the DA dip amplitudes.

The current study on extracellular DA dynamics complements previous works on DA neuron firing for quantitative representation of negative RPE (***Hamid et al., 2016***). Early pioneering works focused on firing of DA neurons (***Schultz et al., 1997; Tobler et al., 2005***), rather than DA itself, to explain variation of the DA dip amplitudes and negative RPEs. Indeed, theoretical works proved that a pause in the DA neuron firing realizes a

continuous representation of both positive and negative RPE (***Bayer et al., 2007***) and a linear modulation in the DA dip amplitudes (***Dreyer et al., 2010***). Experimental studies confirmed that the larger the reward probability, the larger the number of DA neurons that paused their firing upon omission of the reward (***Tian and Uchida, 2015***). These results indicated that firing of DA neurons alone can recapture the DA dip variation. However, unexpected DA release for reward expectations without changes in the firing of DA neurons was found (***Mohebi et al., 2019***). Moreover, presynaptic modulation of the striatal DA release by cholinergic (***Cachope et al., 2012; Chéramy et al., 1986; Threlfell et al., 2012; Zhou et al., 2001***) or glutamatergic (***Glowinski et al., 1988***) innervation was reported, implying DA release regulation independent of DA neuron firing. A prior work using FSCV identified a sequential surge and dip of extracellular DA concentration in the rat ventral striatum after a lottery task, demonstrating symmetrical representation of positive and negative RPE, respectively (***Hart et al., 2014***). It is not clear in the study, however, whether the DA dip was an adjunct to the preceding DA surge or not. Ko and Wanat succeeded with FSCV to isolate a DA dip in the ventral striatum following a PME in an operant task using a PR schedule (***Ko and Wanat, 2016***). But modulation of DA dip amplitudes within a trial was unclear in the study. In the present study, we demonstrated a modulation of a DA dip amplitudes by degree of disappointment. Further investigation is necessary to reveal neurophysiological mechanisms that links extracellular tonic and phasic DA dynamics and DA neuron firing.

One limitation in our DA measurements is that spatial pattern of the virus mediated $GRAB_{DA2m}$ expression did not necessarily match to that of DA receptors in the striatum. While it has been reported that half of the major DA receptors in the striatum, D_1 and D_2 receptors, are distributed in the dendrites and 25% in the spine (***Yung et al., 1995***), subcellular localization of $GRAB_{DA2m}$ is unknown in the current study. Therefore, it is not evident whether neurons expressing dopamine receptors can detect the difference in the extracellular DA levels reported here. Even so, considering the recent report that striatal DA receptor-expressing medium spiny neurons (MSNs) can detect a decrease in DA levels for as short as 0.4 seconds (***Iino et al., 2020***), it is highly likely that MSNs are able to distinguish the small differences in tonic and phasic DA levels. Another caveat in the current study is the stability of the virus expression of $GRAB_{DA2m}$. In this study, DA measurements were completed using the same configuration such as excitation light intensity and detector gain for each mouse within 3−4 weeks after the virus injection surgery, indicating stable $GRAB_{DA2m}$ expression during this period.

We investigated the functional relevance of DA dip variations in behavior adjustment during the FR5 schedule using a computational modeling of the task using a RL algorithm (***Bayer and Glimcher, 2005; Collins and Frank, 2014; Hart et al., 2014; Nakahara et al., 2004; Niv et al., 2007***). Amplitudes of the negative TD error upon a PME increased with PME groups in our agent of the model, reproducing the increasing DA dip

amplitudes with PME groups in mice. We implemented a state transition in our agent to introduce possible misunderstanding of lever press numbers by mice. Our best fitted agent of the model for each mouse had positive transition values, supporting the idea that the mice actually misunderstood the current lever press number occasionally. In conclusion, we propose that the degree of disappointment is represented quantitatively by the magnitude of the DA dip in the ventral striatum and that this DA dip variation is critical for the behavioral adjustment during the FR5 schedule.

Materials and Methods

Animals

All animal procedures were conducted in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and approved by the Animal Research Committee of Keio University School of Medicine (approval 14027-(1)). Experiments were performed using three-month-old male C57BL/6JJmsSlc or C57BL/6NCrSlc mice weighing 23–27 g purchased from SLC (Shizuoka, Japan). The mice were housed individually and maintained on a 12-h light/12-h dark schedule, with lights off at 8:00 PM. Their body weights were maintained at 85% of their initial body weight under conditions of food restriction with water ad libitum.

Adeno-Associated Virus Preparation

For adeno-associated virus (AAV) production, we used the pAAV-hSyn-GRAB-DA2m-W plasmid, a gift from Y. Li. We produced the AAV with the PHP. eB serotype, allowing for the efficient transduction of the central nervous system (***Mathiesen et al., 2020***). Its titer was measured as described elsewhere (***Grieger et al., 2006***). In brief, the plasmid, pHelper (Stratagene), and PHP.eB (Addgene, Plasmid #103005) were transfected into HEK293 cells (AAV293, Stratagene). After three days, the cells were harvested, and the AAV was purified twice using iodixanol. The titer for the AAV was estimated via quantitative PCR.

Surgery

A total of eight mice underwent stereotaxic surgery under anesthesia with a ketamine-xylazine mixture (100 and 10 mg/kg, respectively, i.p.). A 8-mm midline incision was made from the area between the eyes to the cerebellum. A craniotomy with a diameter of up to 1.5 mm was created above the right ventrolateral striatum (VLS) (1.1 mm anterior and 1.9 mm lateral to the bregma) using a high-speed drill, and the dura was surgically removed. A total volume of 0.5 μL $GRAB_{DA2m}$ virus (PHP.eB AAV-hSyn-GRAB-DA2m-W, 1.0×10^{14} genome copies/ml) was injected through a pulled glass micropipette into the VLS (−3.5 to −3.7 mm ventral from dura surface) according to the atlas of Paxinos and Franklin (2004). The injection was driven at a flow rate of 100 nL/min by a Nanoliter 2020 Injector (World Precision Instruments, Sarasota, FL). The micropipette was left in place for another 5 minutes to allow for tissue diffusion before

being retracted slowly. Following the $GRAB_{DA2m}$ virus injection, an optical fiber cannula (CFMC14L05, 400 mm in diameter, 0.39 NA; Thorlabs, Newton, NJ) attached to a ceramic ferrule (CF440-10, Thorlabs) and a ferrule mating sleeve (ADAF1-5, Thorlabs) were inserted into the same side of the VLS as the virus injection and cemented in place (−3.4 to −3.6 mm ventral from dura surface). Operant conditioning and data collection were started in more than 10 days after the surgery to allow the mice to recover and the $GRAB_{DA2m}$ proteins to be expressed.

For negative controls for the virus injection or chronic fiber implantation, another group of six mice underwent stereotaxic surgery under anesthesia, and a midline incision was made as described in the paragraph above. For four of these mice, no craniotomy was made, and the skull was cemented. The other two mice received saline injections bilaterally (0.2 μL/site) targeting the VLS through the craniotomies. The skull was cemented.

Fiber Photometry and Signal Extraction

Extracellular DA fluctuations were measured by a fiber photometric system designed by Olympus Corporation, Tokyo, Japan. Extracellular DA fluorescence signals were obtained by illuminating cells expressing $GRAB_{DA2m}$ with a 465 nm LED (8.0 ± 0.1 μW at the fiber tip) and a 405 nm LED (the same power as that of the 465 nm LED). The 405 nm LED was used to correct for movement artifacts. The 465 nm and 405 nm LED light were emitted alternately at 20 Hz (turned on for 24 ms and off for 26 ms), with the timing precisely controlled by a programmable pulse generator (Master-8, A.M.P.I., Jerusalem, ISRAEL). Each excitation light was reflected by a dichroic mirror (DM455CFP; Olympus) and coupled into an optical fiber cable (400 μm in diameter, 2 meters in length, 0.39 NA, M79L01; Thorlabs, Newton, NJ) through a pinhole (400 μm in diameter). The optical fiber cable was connected to the optical fiber cannula of the mice. The fluorescence signal was detected by a photomultiplier tube with a GaAsP photocathode (H10722–210; Hamamatsu Photonics, Shizuoka, Japan) at a wavelength of 525 nm. The fluorescence signal, TTL signals specifying the durations of the 465 or 405 nm LED excitations, and TTL signals from behavioral settings were digitized by a data acquisition module (cDAQ-9178, National Instruments, Austin, TX). The group of digitized signals was simultaneously recorded at a sampling frequency of 1000 Hz by a custom-made LabVIEW program (National Instruments). The fluorescent signal was processed offline. First, each 24-ms excitation period with each LED was identified according to the corresponding TTL signal. One averaged signal value was computed for each excitation period, with two frames at both ends excluded. Each value of 465 nm excitation was paired with a temporally adjacent counterpart of 405 nm excitation and a ratio of the former to the latter was calculated, yielding a processed signal in a ratiometric manner at a frame rate of 20 Hz.

Operant Conditioning

Behavioral training and tests were performed under constant darkness in an aluminum operant chamber (21.6 × 17.6 × 14.0 cm; Med Associates, Fairfax, VT) housed within a sound-attenuating enclosure. The chamber was equipped with two retractable levers (located 2 cm above the floor), one food magazine between the levers on the floor, and a white noise speaker (80 dB) located on the opposite wall. The mice were required to perform a fixed number of actions (lever presses) to attain a reward: one action on a FR1 schedule or five actions on a FR5 schedule. The mice were first trained for 9–13 consecutive sessions (one 60-minute session/day) on the FR1 schedule followed by six sessions on the FR5 schedule. Training on the FR1 schedule was finished when the mice accomplished seven sessions with more than 50 completed trials per session. The entire experimental procedure took 26–32 days, consisting of surgery, recovery, and training. Each trial began with a sound cue that lasted for 5 s, after which the levers were extended. Presses on the lever on the left of the food magazine (reinforced side) were counted, and a reward pellet (20 mg each, Dustless Precision Pellets, Bio-serv, Flemington, NJ) was dispensed to the magazine immediately after the required number of presses was made. The levers were retracted at the same time as the reward delivery. In contrast, presses on the other lever on the right side had no programmed consequence (non-reinforced side). A refractory period of 0.5 s followed each lever press before the lever was re-extended. In the FR5 sessions, the mice occasionally poked into the magazine before making the required number of lever presses; this was called premature magazine entry (PME). In both PMEs and MEs, the timing of entry was defined as the time point when the distance of the animal's head to the center of the magazine became less than 2.5 cm. A 30-s inter-trial interval followed each food delivery, during which the levers were not presented and the mice consumed the reward. The subsequent trial was automatically initiated after the interval period ended. TTL signals were generated at the timings of the sound cue, lever extension, and lever press and digitized by a data acquisition module (cDAQ-9178, National Instruments, Austin, TX). The TTL signals were simultaneously recorded at a sampling frequency of 1000 Hz by a custom-made LabVIEW program (National Instruments). Both the FR1 and FR5 sessions lasted for 60 min or until the mice received 100 food rewards. Note that, for the control mice, sound cues were not presented. The ITI periods were set at 35 or 30 s. To track the moment-to-moment position of the mice, an infrared video camera (ELP 2 Megapixel WEB Camera, OV2710, Ailipu Technology Co., Ltd, Shenzhen, China) was attached to the ceiling of the enclosure. Reflective tapes were attached to the optical fiber protector (1.2 × 1.4 cm) on the head of the mice. The tapes were recorded at a sampling rate of 20 Hz. The mice's position in each frame was computed offline by a custom-made MATLAB code.

Histology

The mice were anesthetized with an overdose of ketamine (100 mg/kg) and xylazine (10 mg/kg) and intracardially perfused with 4% paraformaldehyde phosphate-buffer solution and decapitated. For a better identification of the fiber track, the optical fiber was not removed from the brain until more than 12 h after perfusion. Subsequently, the brains were cryoprotected in 20% sucrose overnight, frozen, and cut into 50-μm thick sections on a cryostat (Leica CM3050 S, Leica Biosystems, Wetzlar, Germany). The sections were mounted on silane-coated glass slides (S9226, Matsunami Glass, Osaka, Japan). The $GRAB_{DA2m}$ signals received no further amplification. Fluorescence images were captured by an all-in-one microscope (BZ-X710, Keyence, Osaka, Japan).

Data Analyses

A processed ratiometric signal trace was first high-pass filtered at about 0.0167 Hz, corresponding to a wavelength of one minute to exclude low-frequency oscillations. We calculated the z-scores of the DA signal using all inter-trial interval (ITI) periods of a session. Specifically, we first calculated the mean of the DA signals during the last half of the ITI period for each trial and then obtained their mean and standard deviation (SD) for z-score calculation. The reason for only using the last half of the ITI period is that the first half of an ITI period may contain a feeding period. The reason for using all ITI periods of a session for z-score calculation is that there was a huge trial-by-trial variation in the DA signals during the ITI periods within a session. The variation in the SD of the DA signals during the ITI periods *within* the first and second half of a session, respectively, was significantly larger than the variation *between* the first and second half of a session (***Fig. S3H***), indicating that the DA signals during the last half of the ITI periods fluctuated slightly throughout the entire session but heavily fluctuated on an individual trial-by-trial basis. We assume that the large fluctuation among the trials was caused by spontaneous activities of the mice, such as locomotion and grooming, during the ITI period. We confirmed that the DA dip amplitudes calculated with the z-scores were stable during each session (***Fig. S4-3***). For the generation of peri-event plots for the sound cue, lever extension, and ME (***Figure 3***), signals were binned temporally into blocks of 100 ms. For each trial, the binned signal was subtracted by the average signal value of the baseline period (–1 to 0 s relative to the sound cue). In ***Figure 4***, the baseline period was set as 0–0.2 s following the PME or ME. The response amplitude to an event was defined as the mean within the following window: 0–1 s (sound cue/lever extension) or 0.5 to 1.5 s (PME/ME).

Explained Variance

To compare the contributions of the PME presence, session number, and press number to the inter-press-intervals in each FR5 session, we computed the proportions of the variance explained for each factor. For example, the explained variance of the PME presence was calculated as follows:

$$SS_{total} = \sum_{k=1}^{N} \left(Interval_{All}(k) - \overline{Interval_{All}}\right)^2$$

$$SS_{PME} = \left(\overline{Interval_{with\ PME}} - \overline{Interval_{All}}\right)^2 + \left(\overline{Interval_{without\ PME}} - \overline{Interval_{All}}\right)^2$$

$$Explained\ variance_{PME} = {SS_{PME}}/{SS_{total}}$$

where SS_{total} is the sum of the squared differences between each inter-press-interval and the mean of all inter-press intervals; $Interval_{All}(k)$ indicates the duration of the k-th inter-press interval of all N intervals in a given session; and $\overline{Interval_{All}}$ is the average of all interval durations. SS_{PME} specifies the sum of the squared differences between the mean of each condition ($\overline{Interval_{with\ PME}}$ or $\overline{Interval_{without\ PME}}$) and the mean of all inter-press-intervals.

PME Percentage, Overall PME Percentage, and PME Bias

To quantify and compare the number of mice's PMEs across the FR5 sessions, we first computed a value called PME percentage assigned for each inter-press interval group in a session. For example, the PME percentage for the $Interval_{1–2}$, $pPME(Interval_{1–2})$, was calculated as:

$$pPME(Interval_{1\text{-}2}) = 100 \times \frac{nInterval_{1\text{-}2\ with\ PME}}{nInterval_{1\text{-}2}}$$

where $nInterval_{1–2}$ refers to the number of inter-press intervals grouped into $Interval_{1–2}$; $nInterval_{1–2\ with\ PME}$ indicates the number of these intervals with a PME. The other PME percentage values, $pPME(Interval_{2–3})$, $pPME(Interval_{3–4})$ and $pPME(Interval_{4–5})$, were calculated likewise.

We also computed the average of these PME percentage values called overall PME percentage assigned for each session. It was calculated as:

$$Overall\ PME\ percentage = (\ pPME(Interval_{1\text{-}2}) + pPME(Interval_{2\text{-}3}) + pPME(Interval_{3\text{-}4}) + pPME(Interval_{4\text{-}5}))/4$$

To examine the effects of increasing expectations nearer task completion on the mice making PMEs, we then computed a measure called PME bias for each session on the FR5 schedule. It was calculated as:

$$A = pPME(Interval_{3-4}) \times 0.5 + pPME(Interval_{4-5})$$

$$B = pPME(Interval_{1-2}) + pPME(Interval_{2-3}) \times 0.5$$

$$PME\ bias = (A - B)/(A + B)$$

The PME bias is given on a scale from –1 to 1. It is close to 1 (–1) if greater expectations make the mice perform more (fewer) PMEs when they are nearer to task completion (i.e., the number of PMEs during $\text{Interval}_{3\text{–}4}$ or $\text{Interval}_{4\text{–}5}$ is larger [smaller] than that during $\text{Interval}_{1\text{–}2}$ or $\text{Interval}_{2\text{–}3}$). The PME bias is close to 0 if the expectations for task completion have no effect on the number of PMEs (i.e., similar numbers of PMEs are observed during any inter-press intervals).

Simulation of Behavior and Dopamine Dip

At step s, the temporal difference error was computed, and the values were updated as follows:

$$TD\ error(s) = r(s) + \ \gamma \times V_{state}(s+1) - V_{state}(s)$$

$$V_{state}(s) = V_{state}(s) + \ \alpha_critic\ \times TD\ error(s)$$

$$V_{policy}(s,a) = \ V_{policy}(s,a)\ +\ \ \alpha_actor\ \times TD\ error(s)$$

where γ corresponds to the discount factor, and α_critic and α_actor are the learning rates with the state and action, respectively. Suitable parameter values for each mouse' behavior result were determined via a grid search (α_critic = [0.0001, 0.001, 0.01, 0.1], α_actor = [0.0001, 0.001, 0.01, 0.1], γ = [0.6, 0.7, 0.8, 0.9, 1]). The initial state value and action values (press or PME) were chosen from [0, 1, 2, 3, 4, 5, 6, 7, 8, 9], respectively.

After each lever press, the agents selected their next action—either a press or a PME—according to the mice's actual selections. We computed a simulated action selection probability for each decision point, taking the probability of making a $\text{PME}_{3\text{–}4}$ as an example, as follows:

$$f_{PME3-4} \ = \ e^{\left(\beta \frac{V_{PME3-4}}{T}\right)} \Big/ \left(e^{\left(\beta \frac{V_{Press4}}{T}\right)} + \ e^{\left(\beta \frac{V_{PME3-4}}{T}\right)} \right)$$

where β is a constant (chosen from [0.5, 1, 2]), T corresponds to the inverse temperature (chosen from [0.5, 1, 2]), and $V_{PME3\text{–}4}$ and V_{Press4} correspond to the action values for a PME and a press in inter-press interval group 3–4 in a given trial, respectively.

In terms of state transition, we introduced sigma values (chosen from [0, 0.25, 0.5, 1, 2]) for forward and backward transitions. When a lever press was selected, the agents jumps 1, 2, 3, 4, or 5 grids forward according to the probability distribution based on the gaussian kernel (sigma value = original sigma value × number of completed presses in a given trial). The agents were unable to move further beyond the 11th grid—located at the rightmost of our grid world. Any attempt to press at the rightmost grid took the agents to the same 11th grid after the action. Following a PME at the upper row in the grid world, the mice returned (downward shift) to the lower row corresponding to the lever zone. The

mice occasionally jumped backward to one of the lever zone grids located at the lower left of the agents' current position according to the distribution of the gaussian kernel (sigma value = original sigma value × number of completed presses in a given trial). The agents were unable to move further beyond the start grid located at the leftmost of our grid world. Any such attempt took the agents to the starting zone.

The log likelihood (LLH) of a model was computed as follows.

$$LLH = \frac{1}{4N} \sum_{tr=1}^{N} \sum_{IPI=1}^{4} \log_e f(tr, IPI)$$

Where N corresponds to the number of completed trials in the selected FR5 session of each mouse; tr stands for the trial number; and IPI stands for the inter-press interval group number. $f(tr, IPI)$ represents the probability of the action a mouse selects either a press or PME for an inter-press interval group, IPI, in trial tr. We defined the best-fitted model as the model with the greatest LLH for each mouse. The non-reinforced lever presses were not taken into account, as the percentage of these was at a negligible level (***Fig. 1F***).

Statistical Analyses

We analyzed all data using custom codes written in MATLAB and showed the data as the mean ± standard error of the mean (SEM) unless otherwise described. Sample sizes, which were not predetermined but are similar to our previous reports (***Natsubori et al., 2017; Yoshida et al., 2019***), are reported in the Results section and figure legends. To determine the statistical significance, we performed one-sample *t*-tests for a single data set (***Figs. S5E and S4-5A***), one-sample t-tests with Bonferroni corrections for individual data in two or more sample data sets (***Figs. S3B***), paired t-tests for two-sample data sets (***Figs. S2-2C, S2-2D, S2-2G, S4-1B, S4-3A, and S4-5B***), paired t-tests with Bonferroni corrections for multiple two-sample data sets (***Figs. S2-2B, S4-1K, S4-3B, and S4-3C***), and Tukey-Kramer post-hoc analyses for three or more sample data sets (***Figs. 1D, 2E, 3F, 3H, 4D, 4E, S1B, S3H, and S4-1L***). We also employed Kolmogorov-Smirnov tests (***Fig. 2D, S3D, S3E, S3F, and S3G***) and Wilcoxon signed rank-sum tests (***Fig. 5F, S4-1E, and S4-1G***) to compare one or two data distributions. We evaluated the statistical significance of variance among four or more sample data sets using one-way repeated-measures ANOVA (***Figs. 1E, 2H, 2I, 3G, 4G, 4J, 4K, 5I, S1A, S1C, S2-1B, S2-1C, S2-1D, S2-2A, S2-2E, S4-1I, S4-1J, S4-1K, S4-3D, S4-3D, S4-4E, S4-4F, S4-5C, and S5F),*** one-way factorial ANOVA (***Figs. 4F and S4-2A***), and two-way factorial ANOVA (***Figs. 2G, 5H, and S2-1G***). We computed Pearson correlation coefficients to measure linear correlations between the two parameters (***Figs. 4L, 5J, S2-2F, S4-4A, and S4-4C***). The null hypothesis was rejected when $p < 0.05$.

Data and code availability

The original data (behavior and GRAB signals), source data, and TD model codes are provided on Mendeley Data (http://dx.doi.org/10.17632/4g2yg2cxft.2).

Author Contributions

Y.S., N.T., and K.F.T. designed the research. K.F.T. supervised the project. S.Y. created the $GRAB_{DA2m}$ virus. Y.S. and N.T. established the $GRAB_{DA}$ fiber photometry system. Y.S. conducted the fiber photometry experiments, wrote the main manuscript text, and prepared the figures. N.T. edited the main manuscript text.

Acknowledgements

This work was funded by Grant-in-Aid for Brain Mapping by Integrated Neurotechnologies for Disease Studies (Brain/MINDS) JP21dm0207069 from the Agency for Medical Research and Development (AMED), Japan (K.F.T. and S.Y.); JSPS KAKENHI Grant Numbers 19K06944 (N.T.) and 19J01068 (Y.S.).

Competing Interests

The authors declare no competing interests.

References

Amo R, Yamanaka A, Tanaka KF, Uchida N, Watabe-Uchida M. 2020. A gradual backward shift of dopamine responses during associative learning. *bioRxiv* **2020.10.04.325324**. Digital Object Identifier: 10.1101/2020.10.04.325324, PMID: N/A.

Bayer HM, Glimcher PW. 2005. Midbrain Dopamine Neurons Encode a Quantitative Reward Prediction Error Signal. *Neuron* **47**:129-141. Digital Object Identifier: 10.1016/j.neuron.2005.05.020, PMID: 15996553

Bayer HM, Lau B, Glimcher PW. 2007. Statistics of Midbrain Dopamine Neuron Spike Trains in the Awake Primate. *Journal of Neurophysiology* **98**:1428-1439. Digital Object Identifier: 10.1152/jn.01140.2006, PMID: 17615124

Cachope R, Mateo Y, Mathur Brian N, Irving J, Wang H-L, Morales M, Lovinger David M, Cheer Joseph F. 2012. Selective Activation of Cholinergic Interneurons Enhances Accumbal Phasic Dopamine Release: Setting the Tone for Reward Processing. *Cell Reports* **2**:33-41. Digital Object Identifier: 10.1016/j.celrep.2012.05.011 PMID - 22840394, PMID: N/A.

Chéramy A, Romo R, Godeheu G, Baruch P, Glowinski J. 1986. In vivo presynaptic control of dopamine release in the cat caudate nucleus—II. Facilitatory or inhibitory influence ofl-glutamate. *Neuroscience* **19**:1081-1090. Digital Object Identifier: 10.1016/0306-4522(86)90124-7 PMID - 2881227, PMID: N/A.

Collins AGE, Frank MJ. 2014. Opponent Actor Learning (OpAL): Modeling Interactive Effects of Striatal Dopamine on Reinforcement Learning and Choice Incentive. *Psychological Review* **121**:337-366. Digital Object Identifier: 10.1037/a0037015, PMID: 25090423

Covey DP, Cheer JF. 2019. Accumbal Dopamine Release Tracks the Expectation of Dopamine Neuron-Mediated Reinforcement. *Cell Reports* **27**:481-490.e483. Digital Object Identifier: 10.1016/j.celrep.2019.03.055, PMID: 30970251

Dabney W, Kurth-Nelson Z, Uchida N, Starkweather CK, Hassabis D, Munos R, Botvinick M. 2020. A distributional code for value in dopamine-based reinforcement learning. *Nature* **577**:671-675. Digital Object Identifier: 10.1038/s41586-019-1924-6, PMID: 31942076

Dijk WWv, Pligt Jvd, Zeelenberg M. 1999. Effort Invested in Vain: The Impact of Effort on the Intensity of Disappointment and Regret. *Motivation and Emotion* **23**:203-220. Digital Object Identifier: 10.1023/a:1021315314979, PMID: N/A.

Dreyer JK, Herrik KF, Berg RW, Hounsgaard JD. 2010. Influence of Phasic and Tonic Dopamine Release on Receptor Activation. *The Journal of Neuroscience* **30**:14273-14283. Digital Object Identifier: 10.1523/jneurosci.1894-10.2010, PMID: 20962248

Glowinski J, Chéramy A, Romo R, Barbeito L. 1988. Presynaptic regulation of dopaminergic transmission in the striatum. *Cellular and Molecular Neurobiology* **8**:7-17. Digital Object Identifier: 10.1007/bf00712906 PMID - 2900072, PMID: N/A.

Grieger JC, Choi VW, Samulski RJ. 2006. Production and characterization of adeno-associated viral vectors. *Nature Protocols* **1**:1412-1428. Digital Object Identifier: 10.1038/nprot.2006.207 PMID - 17406430, PMID: N/A.

Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, Schmidt R, Weele CM, Kennedy RT, Aragona BJ, Berke JD. 2016. Mesolimbic dopamine signals the value of work. *Nature Neuroscience* **19**:117-126. Digital Object Identifier: 10.1038/nn.4173, PMID: 26595651

Hart AS, Rutledge RB, Glimcher PW, Phillips PEM. 2014. Phasic Dopamine Release in the Rat Nucleus Accumbens Symmetrically Encodes a Reward Prediction Error Term. *The Journal of Neuroscience* **34**:698-704. Digital Object Identifier: 10.1523/jneurosci.2489-13.2014, PMID: 24431428

Heien M, Khan AS, Ariansen JL, Cheer JF, Phillips PEM, Wassum KM, Wightman MR. 2005. Real-time measurement of dopamine fluctuations after cocaine in the brain of behaving rats. *Proceedings of the National Academy of Sciences of the United States of America* **102**:10023-10028. Digital Object Identifier: 10.1073/pnas.0504657102 PMID: 16006505

Howe MW, Dombeck DA. 2016. Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature* **535**:505-510. Digital Object Identifier: 10.1038/nature18942, PMID: 27398617

Iino Y, Sawada T, Yamaguchi K, Tajiri M, Ishii S, Kasai H, Yagishita S. 2020. Dopamine D2 receptors in discrimination learning and spine enlargement. *Nature* **579**:555-560. Digital Object Identifier: 10.1038/s41586-020-2115-1, PMID: 32214250

Kahneman D, Tversky A. 1979. On the interpretation of intuitive probability: A reply to Jonathan Cohen. *Cognition* **7**:409-411. Digital Object Identifier: 10.1016/0010-0277(79)90024-6, PMID: N/A.

Ko D, Wanat MJ. 2016. Phasic Dopamine Transmission Reflects Initiation Vigor and Exerted Effort in an Action- and Region-Specific Manner. *The Journal of Neuroscience* **36**:2202-2211. Digital Object Identifier: 10.1523/jneurosci.1279-15.2016, PMID: 26888930

Mathiesen SN, Lock JL, Schoderboeck L, Abraham WC, Hughes SM. 2020. CNS Transduction Benefits of AAV-PHP.eB over AAV9 Are Dependent on Administration Route and Mouse Strain. *Molecular Therapy - Methods & Clinical Development* **19**:447-458. Digital Object Identifier: 10.1016/j.omtm.2020.10.011 PMID - 33294493, PMID: N/A.

Mohebi A, Pettibone JR, Hamid AA, Wong J-MT, Vinson LT, Patriarchi T, Tian L, Kennedy RT, Berke JD. 2019. Dissociable dopamine dynamics for learning and

motivation. *Nature* **570**:65–70. Digital Object Identifier: 10.1038/s41586-019-1235-y, PMID: 31118513

Moutoussis M, Bentall RP, Williams J, Dayan P. 2009. A temporal difference account of avoidance learning. *Network: Computation in Neural Systems* **19**:137-160. Digital Object Identifier: 10.1080/09548980802192784, PMID: 18569725

Nakahara H, Itoh H, Kawagoe R, Takikawa Y, Hikosaka O. 2004. Dopamine Neurons Can Represent Context-Dependent Prediction Error. *Neuron* **41**:269-280. Digital Object Identifier: 10.1016/s0896-6273(03)00869-9, PMID: 14741107

Natsubori A, Tsutsui-Kimura I, Nishida H, Bouchekioua Y, Sekiya H, Uchigashima M, Watanabe M, d'Exaerde Ad, Mimura M, Takata N, Tanaka KF. 2017. Ventrolateral Striatal Medium Spiny Neurons Positively Regulate Food-Incentive, Goal-Directed Behavior Independently of D1 and D2 Selectivity. *The Journal of Neuroscience* **37**:2723-2733. Digital Object Identifier: 10.1523/jneurosci.3377-16.2017, PMID: 28167674

Neumann Jv, Morgenstern O. 1945. Theory of Games and Economic Behavior. *The Journal of Philosophy* **42**:550-554. Digital Object Identifier: 10.2307/2019327, PMID: N/A.

Niv Y, Daw ND, Joel D, Dayan P. 2007. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology* **191**:507-520. Digital Object Identifier: 10.1007/s00213-006-0502-4, PMID: 17031711

O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. 2003. Temporal Difference Models and Reward-Related Learning in the Human Brain. *Neuron* **38**:329-337. Digital Object Identifier: 10.1016/s0896-6273(03)00169-7, PMID: 12718865

Phillips PEM, Stuber GD, Heien MLAV, Wightman RM, Carelli RM. 2003. Subsecond dopamine release promotes cocaine seeking. *Nature* **422**:614-618. Digital Object Identifier: 10.1038/nature01476, PMID: 12687000

Robinson DL, Venton BJ, Heien MLAV, Wightman RM. 2003. Detecting Subsecond Dopamine Release with Fast-Scan Cyclic Voltammetry in Vivo. *Clinical Chemistry* **49**:1763-1773. Digital Object Identifier: 10.1373/49.10.1763, PMID: 14500617

Rodeberg NT, Sandberg SG, Johnson JA, Phillips PEM, Wightman MR. 2017. Hitchhiker's Guide to Voltammetry: Acute and Chronic Electrodes for in Vivo Fast-Scan Cyclic Voltammetry. *ACS Chemical Neuroscience* **8**:221-234. Digital Object Identifier: 10.1021/acschemneuro.6b00393, PMID: 28127962

Schultz W, Dayan P, Montague RP. 1997. A Neural Substrate of Prediction and Reward. *Science* **275**:1593-1599. Digital Object Identifier: 10.1126/science.275.5306.1593, PMID: 9054347

Starkweather C, Gershman SJ, Uchida N. 2018. The Medial Prefrontal Cortex Shapes Dopamine Reward Prediction Errors under State Uncertainty. *Neuron* **98**:616-629.e616. Digital Object Identifier: 10.1016/j.neuron.2018.03.036, PMID: 29656872

Starkweather CK, Uchida N. 2021. Dopamine signals as temporal difference errors: recent advances. *Current Opinion in Neurobiology* **67**:95-105. Digital Object Identifier: 10.1016/j.conb.2020.08.014, PMID: 33186815

Steiner AP, Redish DA. 2014. Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. *Nature Neuroscience* **17**:995-1002. Digital Object Identifier: 10.1038/nn.3740, PMID: 24908102

Sun F, Zeng J, Jing M, Zhou J, Feng J, Owen SF, Luo Y, Li F, Wang H, Yamaguchi T, Yong Z, Gao Y, Peng W, Wang L, Zhang S, Du J, Lin D, Xu M, Kreitzer AC, Cui G, Li Y. 2018. A Genetically Encoded Fluorescent Sensor Enables Rapid and Specific Detection of Dopamine in Flies, Fish, and Mice. *Cell* **174**:481-496.e419. Digital Object Identifier: 10.1016/j.cell.2018.06.042, PMID: 30007419

Sun F, Zhou J, Dai B, Qian T, Zeng J, Li X, Zhuo Y, Zhang Y, Tan K, Feng J, Dong H, Qian C, Lin D, Cui G, Li Y. 2020. New and improved GRAB fluorescent sensors for monitoring dopaminergic activity in vivo. *bioRxiv* **2020.03.28.013722**. Digital Object Identifier: 10.1101/2020.03.28.013722, PMID: N/A.

Suri RE, Schultz W. 2001. Temporal Difference Model Reproduces Anticipatory Neural Activity. *Neural Computation* **13**:841-862. Digital Object Identifier: 10.1162/089976601300014376, PMID: 11255572

Threlfell S, Lalic T, Platt Nicola J, Jennings Katie A, Deisseroth K, Cragg Stephanie J. 2012. Striatal Dopamine Release Is Triggered by Synchronized Activity in Cholinergic Interneurons. *Neuron* **75**:58-64. Digital Object Identifier: 10.1016/j.neuron.2012.04.038 PMID - 22794260, PMID: N/A.

Tian J, Uchida N. 2015. Habenula Lesions Reveal that Multiple Mechanisms Underlie Dopamine Prediction Errors. *Neuron* **87**:1304-1316. Digital Object Identifier: 10.1016/j.neuron.2015.08.028, PMID: 26365765

Tobler PN, Fiorillo CD, Schultz W. 2005. Adaptive Coding of Reward Value by Dopamine Neurons. *Science* **307**:1642-1645. Digital Object Identifier: 10.1126/science.1105370, PMID: 15761155

Tsutsui-Kimura I, Natsubori A, Mori M, Kobayashi K, Drew MR, d'Exaerde AdK, Mimura M, Tanaka KF. 2017a. Distinct Roles of Ventromedial versus Ventrolateral Striatal Medium Spiny Neurons in Reward-Oriented Behavior. *Current Biology* **27**:3042-3048.e3044. Digital Object Identifier: 10.1016/j.cub.2017.08.061, PMID: 28966085

Tsutsui-Kimura I, Takiue H, Yoshida K, Xu M, Yano R, Ohta H, Nishida H, Bouchekioua Y, Okano H, Uchigashima M, Watanabe M, Takata N, Drew MR, Sano H, Mimura M, Tanaka KF. 2017b. Dysfunction of ventrolateral striatal dopamine receptor type 2-expressing medium spiny neurons impairs instrumental motivation. *Nature Communications* **8**:14304. Digital Object Identifier: 10.1038/ncomms14304, PMID: 28145402

Vasquez TES, McAuley RJ, Gupta NS, Koshy S, Marmol-Contreras Y, Green TA. 2021. Lever-press duration as a measure of frustration in sucrose and drug reinforcement. *Psychopharmacology* **238**:959-968. Digital Object Identifier: 10.1007/s00213-020-05742-2 PMID - 33420805, PMID: N/A.

Yoshida K, Drew MR, Mimura M, Tanaka KF. 2019. Serotonin-mediated inhibition of ventral hippocampus is required for sustained goal-directed behavior. *Nature Neuroscience* **22**:770–777. Digital Object Identifier: 10.1038/s41593-019-0376-5, PMID: 30988523

Yoshida K, Tsutsui-Kimura I, Kono A, Yamanaka A, Kobayashi K, Watanabe M, Mimura M, Tanaka KF. 2020. Opposing Ventral Striatal Medium Spiny Neuron Activities Shaped by Striatal Parvalbumin-Expressing Interneurons during Goal-Directed Behaviors. *Cell Reports* **31**:107829. Digital Object Identifier: 10.1016/j.celrep.2020.107829 PMID - 32610130, PMID: N/A.

Zeelenberg M, Dijk WWv, S.R.Manstead A, Pligt Jd. 2010. The Experience of Regret and Disappointment. *Cognition and Emotion* **12**:221-230. Digital Object Identifier: 10.1080/026999398379727, PMID: N/A.

Zhou F-M, Liang Y, Dani JA. 2001. Endogenous nicotinic cholinergic activity regulates dopamine release in the striatum. *Nature Neuroscience* **4**:1224-1229. Digital Object Identifier: 10.1038/nn769 PMID - 11713470, PMID: N/A.

Supplementary Figures

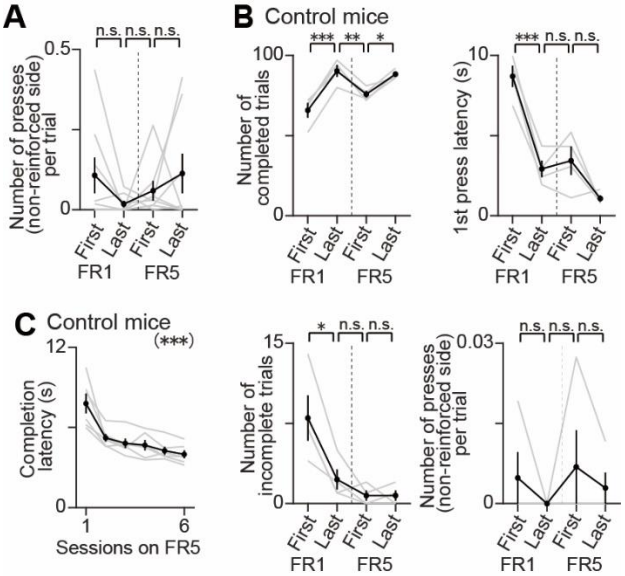

Figure 1—figure supplement 1. The number of non-reinforced lever presses showed no significant changes in a session-by-session comparison.

(A) Session-by-session comparison of the number of presses on the other side, which was not reinforced in the task, per trial among four selected sessions: FR1-first, FR1-last, FR5-first, and FR5-last. The schedule transition from the FR1 to the FR5 schedule happened between FR1-last and FR5-first, as indicated by the vertical dashed line (n = 4 sessions with 8 mice each, Tukey-Kramer post-hoc analysis, FR1-First versus FR1-Last, p = 0.43; FR1-Last versus FR5-First, p = 0.89; FR5-First versus FR5-last, p = 0.79). Gray line, individual mouse. Black line, mean ± SEM.

(B) Session-by-session comparison of the behavior of the control mice, related to ***Figs. 1D and S1A***. The control mice showed similar characteristic behavior changes as the $GRAB_{DA2m}$-expressing mice in terms of the number of completed trials, first lever press latency, number of incomplete trials, and number of non-reinforced lever presses among the four selected sessions around the schedule transition (vertical dashed line) (n = 4 sessions with 8 mice each, Tukey-Kramer post-hoc analysis, FR1-First versus FR1-Last, $p = 2.4 \times 10^{-4}$ [number of completed trials], $p = 1.7 \times 10^{-4}$ [first press latency], $p = 1.1 \times 10^{-2}$ [number of incomplete trials], p = 0.66 [number of presses on unreinforced side]; FR1-Last versus FR5-First, $p = 9.3 \times 10^{-3}$ [number of completed trials], p = 0.91 [first press latency], p = 0.70 [number of incomplete trials], p = 0.39 [number of presses on unreinforced side]; FR5-First versus FR5-last, $p = 2.2 \times 10^{-2}$ [number of completed trials], $p = 5.3 \times 10^{-2}$ [first press latency], p > 0.99 [number of incomplete trials], p = 0.78 [number of presses on unreinforced side]). Gray line, individual mouse. Black line, mean ± SEM. * $p < 0.05$. ** $p < 10^{-2}$. *** $p < 10^{-3}$.

(C) Completion latency, which is the duration between the first to last lever press (***Fig. 2B***, right), detected behavior adjustment during FR5 sessions in the control mice as well as in the $GRAB_{DA2m}$-expressing mice (***Fig. 1E***) (n = 6 sessions with 6 mice each, one-way repeated measures ANOVA, $F_{(5,25)} = 24.00$, $p = 8.5 \times 10^{-9}$). Gray line, individual mouse. Black line, mean ± SEM. *** $p < 10^{-8}$.

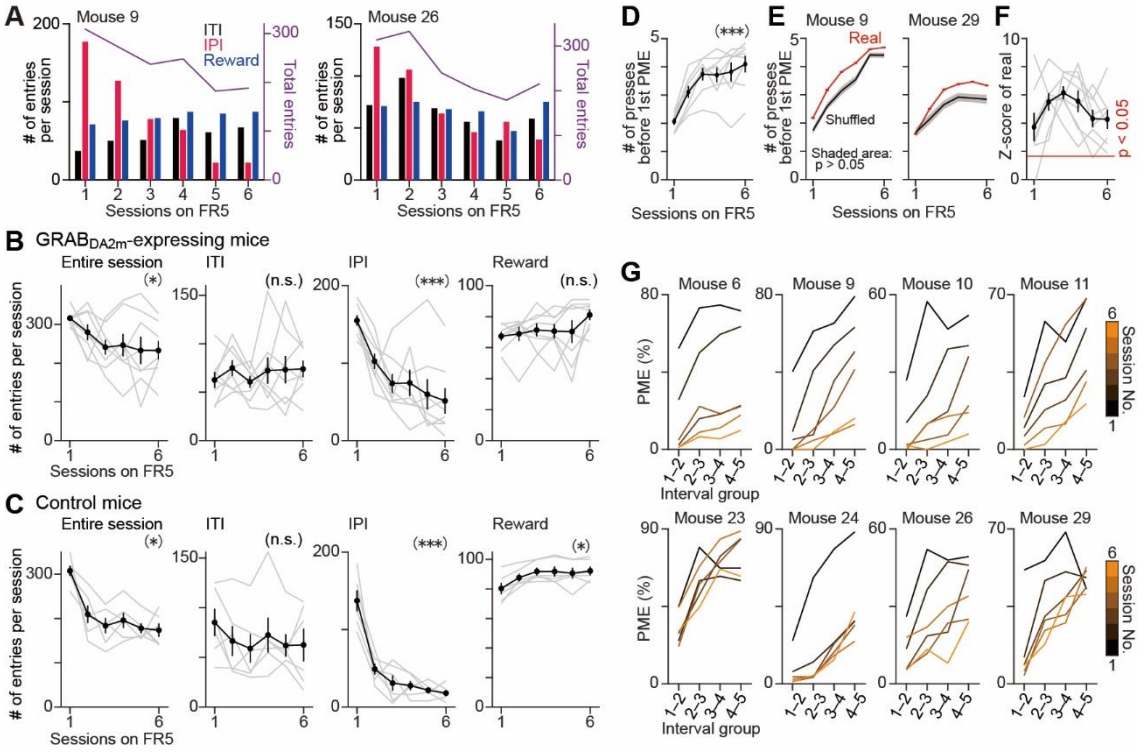

Figure 2—figure supplement 1. Progressive reduction in the total number of head entries across the FR5 sessions is explained by the reduction in the number of PMEs.

(A) Session-by-session comparison of total head entries into the magazine (purple line) and comparisons of the number of head entries during the ITI (black bar), inter-press interval (IPI, red bar), or reward period (blue bar) across the FR5 sessions in two representative mice. Note the gradual decrease in the number of entries during the inter-press interval period corresponding to the change in the total head entries, while there are no such decreases in the number during the other two periods.

(B) Group data showing the session-by-session changes in the number of head entries to the magazine across the FR5 sessions in all mice. Leftmost: The number of total head entries progressively reduced. Second from left: The number of head entries in the ITI periods stayed relatively constant. Second from right: Only the number of head entries during the inter-press intervals significantly reduced, corresponding to the decrease in the total head entries. Rightmost: The number of head entries at the reward periods staying relatively constant (n = 6 sessions with 8 mice each, one-way repeated measures ANOVA, entire session, $F_{(5,35)} = 3.00$, $p = 2.4 \times 10^{-2}$; ITI, $F_{(5,35)} = 0.47$, $p = 0.80$; IPI, $F_{(5,35)} = 11.98$, $p = 8.4 \times 10^{-7}$; reward, $F_{(5,35)} = 1.62$, $p = 0.18$). Gray line, individual session. Black line, mean ± SEM. * $p < 0.05$. *** $p < 10^{-6}$.

(C) Group data showing the session-by-session changes in the number of head entries to the magazine across the FR5 sessions in the control mice, related to ***Fig. S2-1B***. Similar to the $GRAB_{DA2m}$-expressing mice, the number of total head entries to the magazine progressively decreased (leftmost panel), and only the entries during the inter-press interval period (second from the right) accounted for this decrease (n = 6 sessions with 8 mice each, one-way repeated measures ANOVA, entire session, $F_{(5,25)} = 16.06$, $p = 4.1 \times 10^{-7}$; ITI, $F_{(5,25)} = 1.58$, $p = 0.20$; IPI,

$F_{(5,25)} = 29.29$, $p = 1.1 \times 10^{-9}$; reward, $F_{(5,25)} = 10.80$, $p = 1.3 \times 10^{-5}$). Gray line, individual session. Black line, mean ± SEM. * $p < 10^{-4}$. *** $p < 10^{-8}$.

(D) The average number of presses before the mice made their first PME or ME in a given trial, a behavioral readout reflecting how many presses the mice expected as adequate for a reward, progressively increased across the FR5 sessions in all mice tested (n = 6 sessions with 8 mice each, one-way repeated measures ANOVA, entire session, $F_{(5,35)} = 29.16$, $p = 1.5 \times 10^{-11}$). *** $p < 10^{-10}$.

(E) The average number of presses before their first PME or a ME in a given trial (red line, mean ± SEM) was significantly larger than those with PME group number-shuffled counterparts (black line and grey shaded data, mean ± 1.65 standard deviation [$p > 0.05$], shuffled 1000 times) in two representative mice, suggesting that the mice recognize the press number or PME group number and organize the timings of their reward-seeking behavior accordingly throughout the FR5 sessions.

(F) Group data of the z-scored number of presses before a first PME or a ME in a given trial—calculated based on the corresponding shuffled datasets—across the FR5 sessions in all mice. Z-scores of 46 sessions out of 48 sessions corresponded to $p < 0.05$ (red line).

(G) Session-by-session plots of the PME percentage versus the interval group during the FR5 schedule of all mice, related to ***Fig. 2G*** (n = 24 data points, 4 interval groups × 6 sessions, two-way factorial ANOVA, interval group number, $F_{(3,23)} = 17.96$, $p = 3.2 \times 10^{-5}$ [mouse 6], $F_{(3,23)} = 18.48$, $p = 2.7 \times 10^{-5}$ [mouse 9], $F_{(3,23)} = 8.49$, $p = 1.6 \times 10^{-3}$ [mouse 10], $F_{(3,23)} = 28.83$, $p = 1.8 \times 10^{-6}$ [mouse 11], $F_{(3,23)} = 28.62$, $p = 1.9 \times 10^{-6}$ [mouse 23], $F_{(3,23)} = 25.34$, $p = 4.0 \times 10^{-6}$ [mouse 24], $F_{(3,23)} = 14.87$, $p = 9.2 \times 10^{-5}$ [mouse 26], $F_{(3,23)} = 11.98$, $p = 2.9 \times 10^{-4}$ [mouse 29]; session number, $F_{(5,23)} = 96.90$, $p = 7.1 \times 10^{-11}$ [mouse 6], $F_{(5,23)} = 25.64$, $p = 7.5 \times 10^{-7}$ [mouse 9], $_{(5,23)} = 19.56$, $p = 4.3 \times 10^{-6}$ [mouse 10], $F_{(5,23)} = 22.88$, $p = 1.6 \times 10^{-6}$ [mouse 11], $F_{(5,23)} = 3.18$, $p = 3.8 \times 10^{-2}$ [mouse 23], $F_{(5,23)} = 24.76$, $p = 9.5 \times 10^{-7}$ [mouse 24], $F_{(5,23)} = 12.95$, $p = 5.3 \times 10^{-5}$ [mouse 26], $F_{(5,23)} = 4.87$, $p = 7.6 \times 10^{-3}$ [mouse 29]).

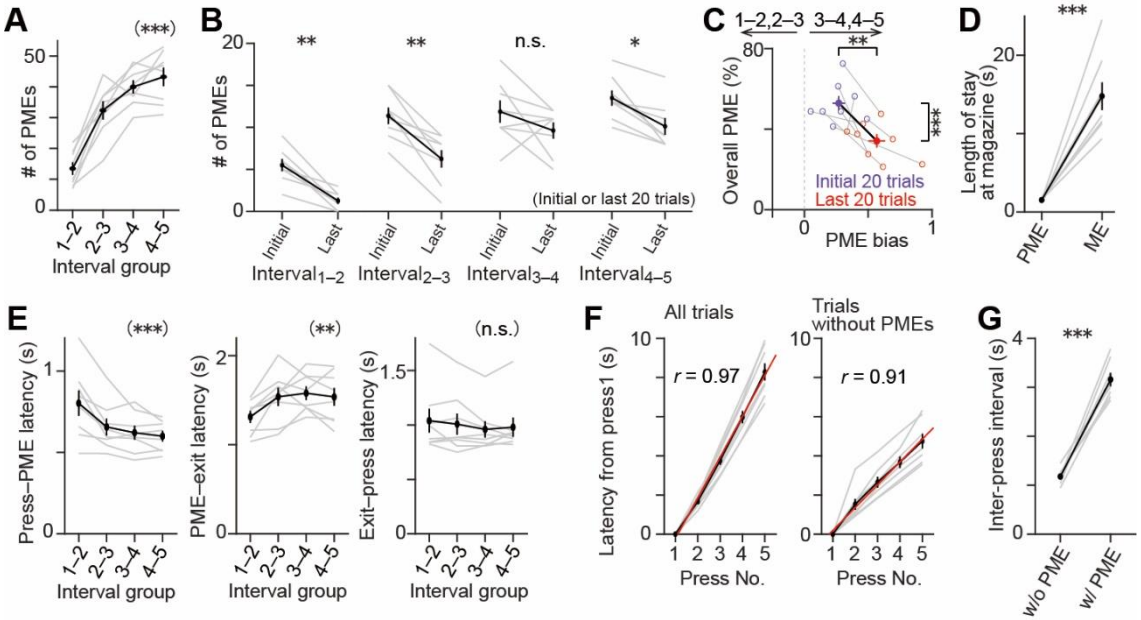

Figure 2—figure supplement 2. Analysis of the single FR5 session from each mouse selected for further behavioral or fiber photometry analyses.

(A) Group data showing that the number of PMEs increased with an increase in the interval group number in the selected FR5 sessions in all mice (n = 4 interval groups with 8 sessions each from 8 mice, number of PMEs = 13.5 ± 2.0 [$\text{Interval}_{2–3}$], 32.4 ± 2.9 [$\text{Interval}_{2–3}$], 40.0 ± 2.2 [$\text{Interval}_{3–4}$], 43.3 ± 3.0 [$\text{Interval}_{4–5}$], one-way repeated measures ANOVA, $F_{(3,21)} = 71.25$, p = 3.6×10^{-11}). The minimum number of events at a PME group was 7. Gray line, individual session. Black line, mean ± SEM. *** p < 10^{-10}.

(B) Group data of a within-session comparison of the number of events in each PME group between the initial and final 20 trials in the selected FR5 sessions showing a consistent decrease in all mice (n = 8 trial groups from 8 mice each, number of PMEs = 5.5 ± 0.7 [$\text{Interval}_{1–2}$, initial trials], 1.3 ± 0.4 [$\text{Interval}_{1–2}$, last trials], 11.4 ± 1.0 [$\text{Interval}_{2–3}$, initial trials], 6.3 ± 1.0 [$\text{Interval}_{2–3}$, last trials], 11.9 ± 1.3 [$\text{Interval}_{3–4}$, initial trials], 9.6 ± 1.0 [$\text{Interval}_{3–4}$, last trials], 13.5 ± 0.9 [$\text{Interval}_{4–5}$, initial trials], 10.1 ± 1.0 [$\text{Interval}_{4–5}$, last trials], paired t-test with Bonferroni correction, $\text{Interval}_{1–2}$, $t_{(7)} = 5.86$, p = 2.5×10^{-3}; $\text{Interval}_{2–3}$, $t_{(7)} = 5.09$, p = 5.7×10^{-3}; $\text{Interval}_{3–4}$, $t_{(7)} = 1.46$, p = 0.75; $\text{Interval}_{4–5}$, $t_{(7)} = 4.22$, p = 0.016). Filled circle, mean ± SEM within a group of trials. Gray line, individual mouse. Black line, mean ± SEM. * p < 0.05.** p < 10^{-2}.

(C) Group data plot showing that the increase in the PME bias and decrease in the overall PME percentage not only happened across sessions (***Fig. 2H***) but also within the selected FR5 sessions (n = 2 trial groups with 8 mice each, PME bias, 0.27 ± 0.05 [initial trials], 0.57 ± 0.07 [last trials], paired t-test, $t_{(7)} = –3.65$, p = 8.2×10^{-3} ; overall PME percentage, 52.8 ± 3.6% [initial trials], 34.1 ± 3.4% [last trials], paired t-test, $t_{(7)} = –5.58$, p = 8.3×10^{-4}). Gray line, individual mouse. Black line, mean of all mice. ** p < 10^{-2}.*** p < 10^{-3}.

(D) The mice stayed longer at the food magazine after an ME than after a PME (n = 2 groups with 8 mice each, PME, 1.53 ± 0.08 s; ME, 14.79 ± 1.76 s; paired t-test, $t_{(7)} = –7.37$, p = 1.5×10^{-4}). Gray line, individual mouse. Black line, mean ± SEM. *** p < 10^{-3}.

(E) Inter-press intervals with a PME were divided into three parts (press–PME, PME–exit, and exit–press periods).

Left: We considered the press–PME latency as related to the measure of frustration—the lever press duration, which is reported to be positively correlated with the level of frustration. We

discovered a decreasing trend for the press-to-PME latency with an increasing PME group number (n = 4 interval groups with 8 sessions each from 8 mice, one-way repeated measures ANOVA, $F_{(3,21)} = 9.82$, $p = 3.0 \times 10^{-4}$), suggesting that the mice were not relatively more frustrated in later presses in an FR5 trial. Center: Significant differences in the length of stay at the magazine following PMEs in the inter-press interval groups, suggesting that it took more time for the mice to realize the same outcome of reward omission with a greater degree of reward expectations, which may correlate with the degree of disappointment (n = 4 interval groups with 8 sessions each from 8 mice, one-way repeated measures ANOVA, $F_{(3,21)} = 5.28$, $p = 7.2 \times 10^{-3}$). Right: No significant differences in the latency of a subsequent press following the preceding magazine exit moment (n = 4 interval groups with 8 sessions each from 8 mice, one-way repeated measures ANOVA, $F_{(3,21)} = 1.56$, $p = 0.23$). Gray line, individual session. Black line, mean ± SEM. ** $p < 10^{-2}$.*** $p < 10^{-3}$.

(F) The lever press number was positively correlated with the elapsed time, since the first press in a given trial (regardless of the trials with PMEs) is included (left panel) or not (right panel) (n = 5 press numbers with 8 sessions each from 8 mice, Pearson correlation coefficient, all trials, $r = 0.97$, $p = 2.0 \times 10^{-25}$; trials without PMEs, $r = 0.91$, $p = 2.7 \times 10^{-16}$).

(G) Presence of a PME significantly elongated the duration of the corresponding inter-press interval (n = 2 groups with 8 sessions each from 8 mice, without PME, 1.18 ± 0.05 s; with PME, 3.16 ± 0.13 s; paired t-test, $t_{(7)} = -12.53$, $p = 4.8 \times 10^{-6}$).

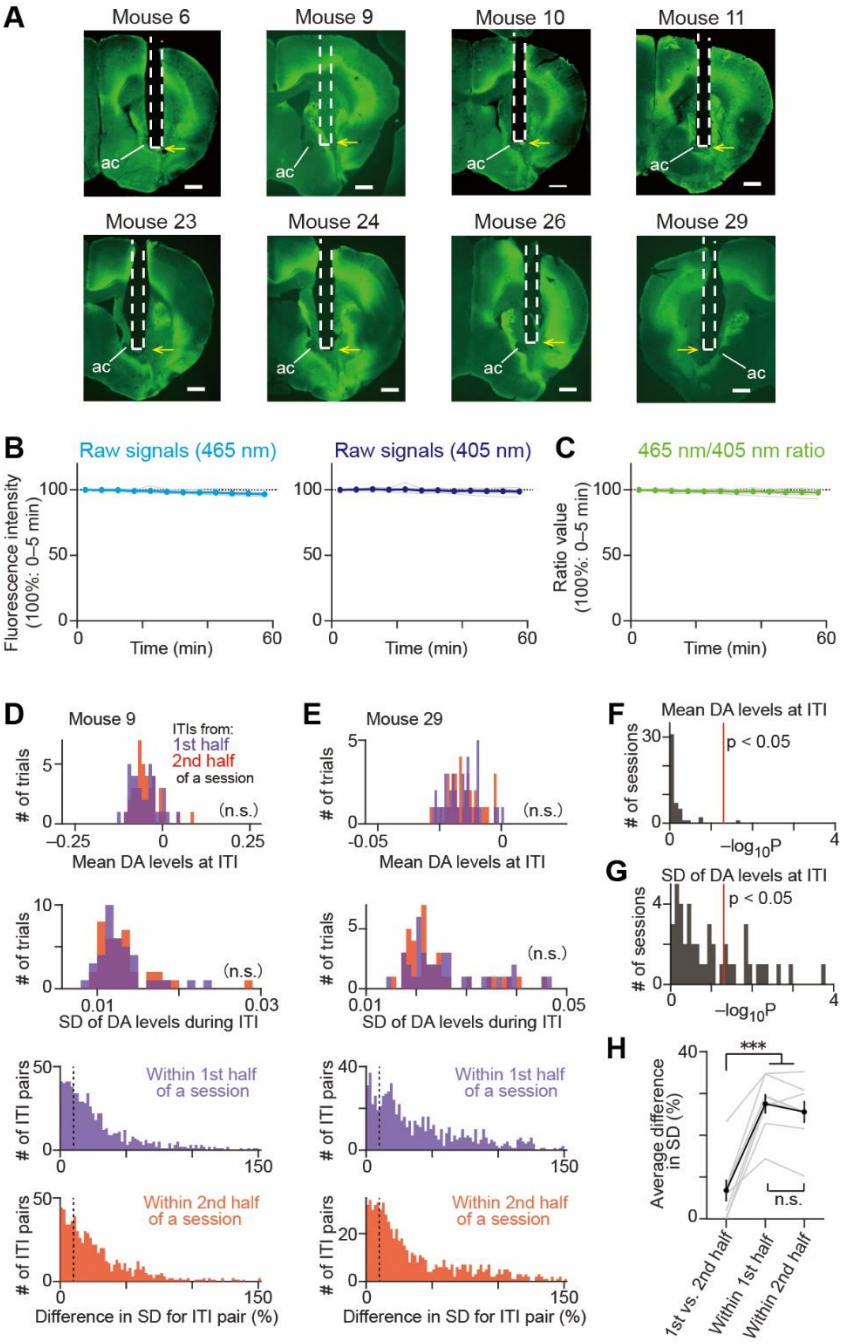

Figure 3—figure supplement 1. Stable fluorescence intensity for an hour enabled stable *in vivo* DA measurements from freely behaving mice.

(A) Coronal brain section showing the location of the optic fiber and the $GRAB_{DA2m}$ expression in the VLS of all mice tested. Dashed line, fiber track. ac: anterior commissure. Yellow arrow, fiber tip.

(B) Fluorescence intensity changes by 405 nm (left) and 465 nm (right) excitation over one hour. We observed no significant reduction in intensity from 100% (0–5 min) due to fluorescence photobleaching for any signals at any periods (n = twelve 5-min bins with 6 sessions each from 3 mice, 2 sessions from each mice, one-sample t-test with Bonferroni correction, 405 nm excitation, fluorescence intensity > 98.8 ± 1.1%, all $t_{(5)} > -1.17$, all p

> 0.99; 465 nm excitation, fluorescence intensity > 96.7 ± 0.8%, all $t_{(5)} > -4.13$, all p > 0.11). Note that the mice were individually placed in an operant chamber but not subjected to a task. Horizontal dotted line, 100% fluorescence intensity (0–5 min). Gray line, individual session. Thicker line with filled circles, mean ± SEM.

(C) Same as (B) but for the 465 nm/405 nm ratio. We observed no significant reduction in ratio from 100% at any periods (n = twelve 5-min bins with 6 sessions each from 3 mice, 2 sessions from each mice, one-sample t-test with Bonferroni correction, ratio > 98.0 ± 1.2%, all $t_{(5)} > -1.75$, all p > 0.99).

(D) Investigation of the progressive reduction in the DA signal fluctuation ranges between the first and second half of the trials of a session in a representative mouse. Top: Comparison of the distributions of the average DA signal levels in the 465 nm/405 nm ratio (high-pass filtered at 0.0167 Hz) in each ITI period between the first and the second half of a session, showing no significant difference (n = 38 trials each, Kolmogorov-Smirnov test, $D_{max} = 0.11$, p = 0.64). Second from top: Same as the top panel but for the comparison of the distributions of the standard deviation, the results of which exclude the possibility of dynamic changes in the standard deviation of the DA release ($D_{max} = 0.11$, p = 0.64). Bottom two panels: Differences in the SD within each half of a session, suggesting large within-section variations in the SD (n = 703 trial pairs from 38 trials, first half, 71.4% of pairs (502 trials) with more than a 10% difference; second half, 72.8% of pairs (512 trials) with more than a. 10% difference). Vertical dashed line, 10% difference.

(E) Same as ***Fig. S2-2D*** but for the investigation of the progressive reduction in the DA signal fluctuation ranges in another representative mouse. Top: Comparison of the distributions of the average DA signal levels (n = 38 and 37 trials for the first and second halves, respectively, Kolmogorov-Smirnov test, $D_{max} = 0.05$, p = 0.90). Second from top: Same as the top panel but for the comparison of the distributions of the SD ($D_{max} = 0.19$, p = 0.22). Bottom two panels: Differences in the SD within each half of a session (n = 703 trial pairs from 38 trials, first half, 80.8% of pairs (568 trials) with more than 10% difference; second half, 75.8% of pairs (505 trials) with more than 10% difference). Vertical dashed line, 110% difference.

(F) In the majority of the sessions, we did not find significant reductions in the mean DA levels at ITI periods between the first and second half of a session (Kolmogorov-Smirnov test, p > 0.05 in 47 out of 48 sessions from 8 mice). Red line: p = 0.05.

(G) In the majority of the sessions, we did not find significant reductions in the SD of the DA levels at ITI periods between the first and second half of a session (Kolmogorov-Smirnov test, p > 0.05 in 30 out of 48 sessions from 8 mice). Red line: p = 0.05.

(H) We found significantly large differences in the SDs of the DA levels in the ITIs within the first and second halves of the sessions compared to the differences caused by a progressive change in the SDs between the first and second sessions. Since the normalization of the DA signals for z-score calculation using SDs varying among the trials rather than across the sections may impair a comparison of the DA response amplitudes throughout a session, these results encouraged us to utilize a uniform normalization factor for the z-score calculation for each session (n = 3 groups with 8 mice each, between halves (first versus second halves), 6.8 ± 2.6%, within first half , 27.5 ± 2.3%, within second half, 25.6 ± 2.6%, Tukey-Kramer post-hoc analysis, between halves vs within first half, $p = 1.6 \times 10^{-6}$; between halves vs within second half, $p = 5.1 \times 10^{-6}$; within 1st half vs within 2nd half, p = 0.71). Gray line, individual mouse. Black line, mean ± SEM.

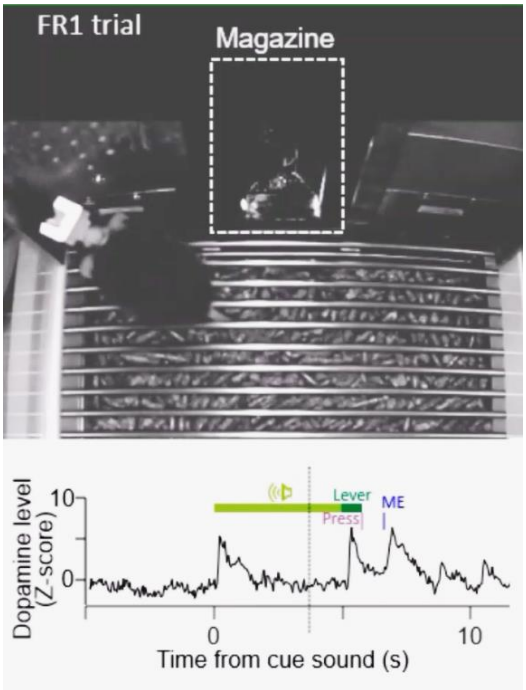

Figure 3—video 1. Mouse behavior and extracellular dopamine fluctuations during the FR1 task.

(Top) Representative behavior of a mouse engaging in the FR1 task. After the 5-s noise sound cue, reinforced (left) and non-reinforced (right) levers were presented. The mouse was rewarded at the magazine (white dashed line) after pressing the reinforced lever once per trial.

(Bottom) Simultaneous measurement of extracellular DA fluctuations in the right ventral striatum. Vertical dotted line indicates behavior timing. A phasic DA increase was observed upon the onset of a sound cue (light green bar), lever extension (dark green bar), or ME (blue tick).

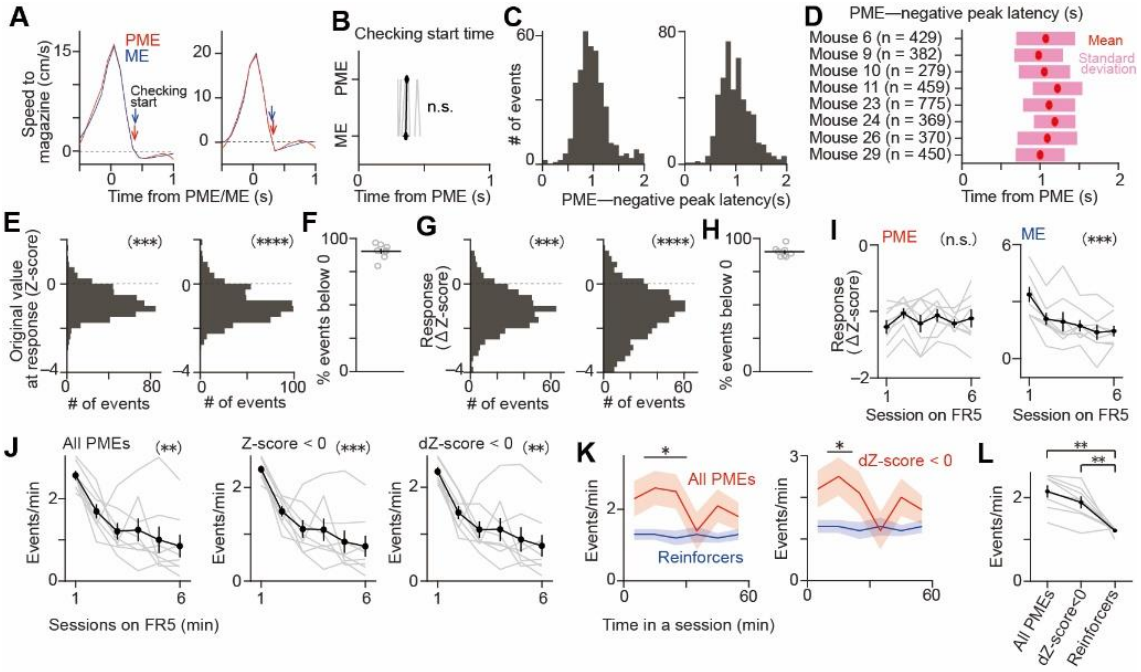

Figure 4—figure supplement 1. PMEs induce DA dips.

(A) Average changes in the head speed toward the magazine around the PME or ME in two representative mice. A positive value corresponds to the mice's approaching the magazine. We defined the putative onset of magazine checking behavior as the time when the head speed became equal to or less than 0 cm/s following the entry. Arrowhead, putative starting timing of magazine checking. Red line and shade, PME. Blue line and shade, ME. Data shown as mean ± SEM. Left panel: Mouse 10, n = 397 PMEs and 403 MEs from 6 FR5 sessions. Right panel: Mouse 23, n = 861 PMEs and 360 MEs from 6 sessions.

(B) Group data showing the putative checking start time after a PME/ME in all mice. According to these results, we set the baseline period for the calculation of the DA responses upon a PME/ME as 0–0.2 s (n = 2 data groups with 8 mice each, PME, 0.37 ± 0.02 s, ME, 0.36 ± 0.02, paired t-test, $t_{(7)} = 1.30$, p = 0.24). Gray line, individual mouse. Black line, mean ± SEM.

(C) Temporal distributions of the negative peaks of DA responses following PMEs in two representative mice. The responses with a large amplitude Δz-score < –2.33 at the peak were selected and analyzed. Left panel: Mouse 9, n = 382 events from 6 sessions, 0.98 ± 0.31 s (mean ± standard deviation). Right panel: Mouse 29, n = 450 events from 6 sessions, 1.00 ± 0.31 s (mean ± standard deviation).

(D) The average latency and standard deviation of a negative peak in the DA responses following PMEs in all mice recorded. According to these results, we set the period for the calculation of a DA response upon a PME as 0.5–1.5 s. Average latency, 1.09 ± 0.03 s (n = 8 mice, mean ± SEM). Filled circle, mean. Shade, standard deviation.

(E) Distributions of the z-scores in the DA response period (0.5–1.5 s following a PME, refer to the ***Fig. S4-1D***) in two representative mice. Left panel: Mouse 9, n = 490 events from 6 sessions, z-score = –0.85 ± 0.03. We found that 90.8% of the responses (445 events) showed z-scores below zero (Wilcoxon signed rank test, signed rank = 3957, Z = –17.92, p = 8.5×10^{-72}). Right panel: Mouse 29, n = 648 events from 6 sessions, z-score = –0.95 ± 0.03. 89.8% (582 events) showed z-scores below zero (signed rank = 9596, Z = –20.04, p = 2.4×10^{-89}). *** p < 10^{-71}. **** p < 10^{-88}.

(F) Group data showing the percentage of DA responses with z-scores below zero in all mice recorded. 90.3 ± 2.0% (n = 8 mice).

(G) Same as ***Fig. S4-1E*** but for distributions of the DA response amplitude (Δz-score) in two representative mice. Left panel: Mouse 9, n = 490 events from 6 sessions, z-score = –1.15 ± 0.04. 90.2% of the responses (442 events) showed negative deflections (Δz-score < 0) (Wilcoxon signed rank test, signed rank = 4646, Z = –17.70, $p = 4.3 \times 10^{-70}$). Right panel: Mouse 29, n = 648 events from 6 sessions, z-score = –1.37 ± 0.05. 88.7% of the responses (575 events) showed negative deflections (signed rank = 8992, Z = –20.17, $p = 1.9 \times 10^{-90}$). *** $p < 10^{-69}$. **** $p < 10^{-89}$.

(H) Group data showing the percentage of DA responses with negative deflections (Δz-score < 0) in all mice recorded. 89.94 ± 1.35% (n = 8 mice).

(I) No significant changes in the response amplitude upon a PME (left panel) but upon a ME (right panel) across the FR5 sessions in all mice recorded (n = 6 sessions with 8 mice each, one-way repeated measures ANOVA, PME, $F_{(5,35)} = 1.68$, p = 0.17; ME, $F_{(5,35)} = 25.42$, $p = 9.7 \times 10^{-11}$), suggesting that a DA response upon a PME in any FR5 session has the potential to serve as, in the RL theory, a teaching signal with a similar weight throughout the FR5 sessions. Gray line, individual mouse. Black line, mean ± SEM. *** $p < 10^{-10}$.

(J) Group data showing progressive reductions in the frequency of all PMEs (left), PMEs followed by a DA response with the z-score < 0 (center), and those followed by a negative deflection (Δz-score < 0) across the FR5 sessions in all mice (n = 6 sessions with 8 mice each, one-way repeated measures ANOVA, all PMEs, $F_{(5,35)} = 11.82$, $p = 9.7 \times 10^{-7}$; PMEs with DA responses with negative z-scores, $F_{(5,35)} = 17.98$, $p = 8.2 \times 10^{-9}$; PMEs with DA responses with negative deflections, $F_{(5,35)} = 13.18$, $p = 3.0 \times 10^{-7}$). Gray line, individual mouse. Black line, mean ± SEM. ** $p < 10^{-6}$. *** $p < 10^{-8}$.

(K) Within-session changes in the frequency of all PMEs (left, red line) and those followed by a negative deflection (right, red line) plotted together with the frequency of task accomplishments (blue line) in the selected FR5 session of a representative mouse. Bar with an asterisk, p < 0.05 (paired t-test with Bonferroni correction).

(L) Group data showing that the frequency of all PMEs and PMEs with DA responses with negative deflections is higher than that of the mice's obtaining reinforcers in the selected FR5 sessions of all mice (n = 3 data groups with 8 mice each, number of events per minute = 2.2 ± 0.2 [all PMEs], 1.9 ± 0.1 [responses with Δz -score < 0], 1.2 ± 0.0 [reinforcers], paired t-test with Bonferroni correction , all PMEs versus reinforcers, $t_{(7)} = 7.72$, $p = 1.4 \times 10^{-3}$, responses with Δz -score < 0 versus reinforcers, $t_{(7)} = 4.04$, $p = 9.8 \times 10^{-3}$). Gray line, individual mouse. Black line, mean ± SEM. ** $p < 10^{-2}$.

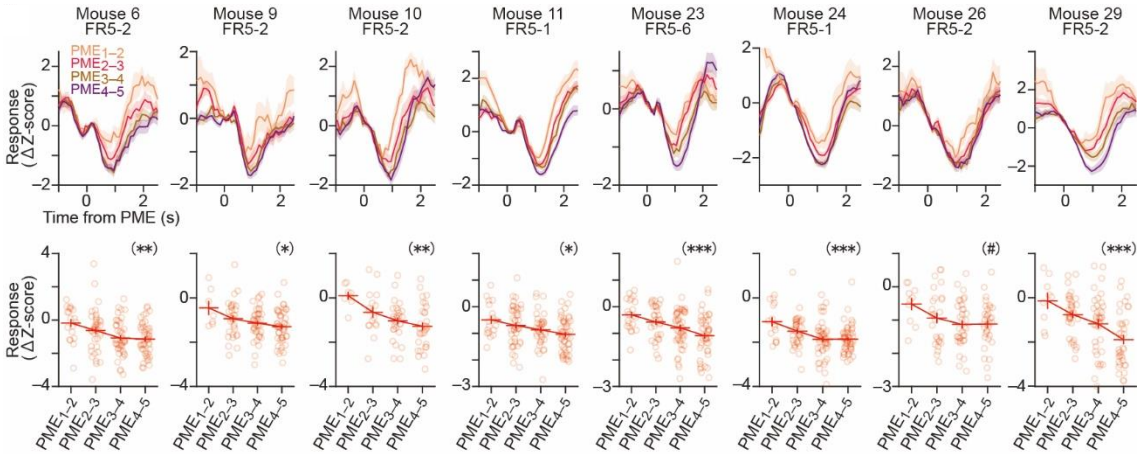

Figure 4—figure supplement 2. Plots of DA dip graduality upon PMEs in all mice recorded.

$GRAB_{DA2m}$ traces (top row) and corresponding statistics (bottom row) of response amplitude for each PME group in the selected FR5 sessions of all the mice recorded (related to ***Fig. 4F***). The magnitude of a DA response following a PME consistently increased with the increasing PME group number (one-way ANOVA, mouse 6, $F_{(3,146)} = 4.61$, $p = 4.1 \times 10^{-3}$; mouse 9, $F_{(3,126)} = 3.28$, $p = 2.3 \times 10^{-2}$; mouse 10, $F_{(3,84)} = 5.66$, $p = 1.4 \times 10^{-3}$; mouse 11, $F_{(3,150)} = 3.87$, $p = 1.1 \times 10^{-2}$; mouse 23, $F_{(3,146)} = 6.79$, $p = 2.6 \times 10^{-4}$; mouse 24, $F_{(3,151)} = 6.26$, $p = 5.0 \times 10^{-4}$; mouse 26, $F_{(3,105)} = 2.28$, $p = 0.084$; mouse 29, $F_{(3,117)} = 8.27$, $p = 5.0 \times 10^{-5}$). # $p < 0.1$. * $p < 0.05$. ** $p < 10^{-2}$. *** $p < 10^{-3}$.

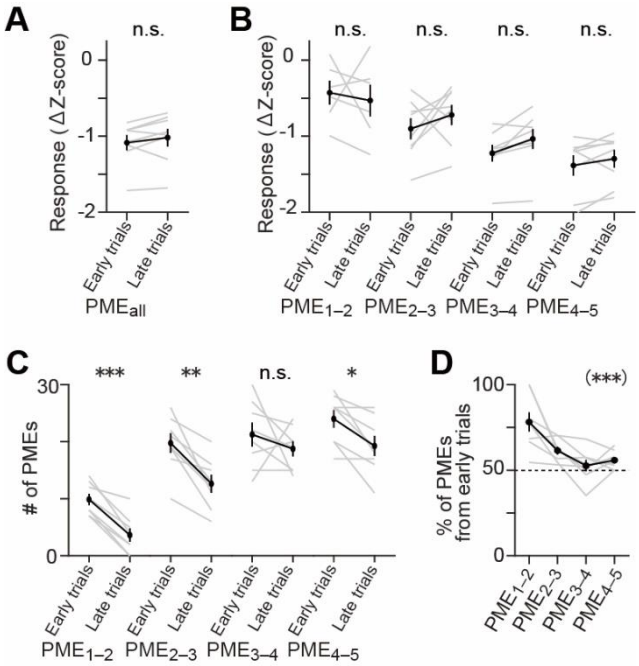

Figure 4—figure supplement 3. No significant within-session progressive changes in DA dip amplitude upon PMEs.

(A–D) Dopaminergic and behavioral analyses excluding the possibility that within-session progressive reductions in the DA responses upon PMEs within a session may explain the variability in the DA dip amplitudes.

(A) Group data showing that the DA response (dip) amplitude was not consistently changed between the early trials (first half) and late trials (second half) in the selected FR5 sessions of all mice. The result contradicted our hypothesis that the amplitude would progressively decrease as the mice became familiar with the task (n = 8 sessions each from 8 mice, response amplitude (Δz-score) = –1.09 ± 0.10 [early trials], –1.02 ± 0.12 [late trials]; paired t-test, $t_{(7)} = -1.45$, p = 0.19). Gray line, individual mouse. Black line, mean ± SEM.

(B) Related to ***Fig. S4-3A*** but for the absence of significant progressive reductions in the DA response amplitude for each PME group, suggesting that such progressive reductions cannot explain the variability in the DA dip amplitude among the PME groups (n = 6–8 sessions each from 6–8 mice, response amplitude (Δz-score) = –0.43 ± 0.16 [$PME_{1–2}$, early trials], –0.53 ± 0.21 [$PME_{1–2}$, late trials], –0.90 ± 0.14 [$PME_{2–3}$, early trials], –0.72 ± 0.13 [$PME_{2–3}$, late trials], –1.22 ± 0.11 [$PME_{3–4}$, early trials], –1.04 ± 0.13 [$PME_{3–4}$, late trials], –1.38 ± 0.13 [$PME_{4–5}$, early trials], –1.3 ± 0.12 [$PME_{4–5}$, late trials], paired t-test with Bonferroni correction, $PME_{1–2}$, $t_{(5)} = 0.42$, p > 0.99; $PME_{2–3}$, $t_{(7)} = -1.08$, p > 0.99; $PME_{3–4}$, $t_{(7)} = -2.94$, p = 0.087; $PME_{4–5}$, $t_{(7)} = -1.32$, p = 0.91). Gray line, individual mouse. Black line, mean ± SEM.

(C) Group data showing that the number of PMEs significantly and consistently decreased within a session for most PME groups in all mice (n = 8 sessions each from 8 mice, number of PMEs = 9.9 ± 1.0 [$PME_{1–2}$, early trials], 3.6 ± 1.2 [$PME_{1–2}$, late trials], 19.8 ± 1.7 [$PME_{2–3}$, early trials], 12.6 ± 1.6 [$PME_{2–3}$, late trials], 21.3 ± 2.1 [$PME_{3–4}$, early trials], 18.8 ± 1.3 [$PME_{3–4}$, late trials], 24.0 ± 1.6 [$PME_{4–5}$, early trials], 19.3 ± 1.8 [$PME_{4–5}$, late trials], paired t-test with Bonferroni correction, $PME_{1–2}$, $t_{(7)} = 7.44$, $p = 5.8 \times 10^{-4}$; $PME_{2–3}$, $t_{(7)} = 4.66$, $p = 9.3 \times 10^{-3}$; $PME_{3–4}$, $t_{(7)} = 0.92$, p > 0.99; $PME_{4–5}$, $t_{(7)} = 3.37$, p > 0.048). Gray line, individual mouse. Black line, mean ± SEM.

(D) Comparison of the proportions of the PMEs occurred in early trials across the PME groups. Data for the smaller PME group numbers consisted of relatively more events occurring in the early trials, which is against our hypothesis that smaller PME group numbers, inducing smaller DA responses, would be more likely in later trials, the part hypothesized to induce smaller DA dips (n = 8 sessions each from 8 mice, percent of PMEs from early trials, 78.2 ± 5.6% [$PME_{1–2}$], 61.5 ± 2.3% [$PME_{2–3}$], 52.6 ± 3.4% [$PME_{3–4}$], 55.8 ± 1.9% [$PME_{4–5}$], one-way repeated-measures ANOVA, $F_{(3,21)} = 11.46$, $p = 1.2 \times 10^{-4}$). Gray line, individual mouse. Black line, mean ± SEM.

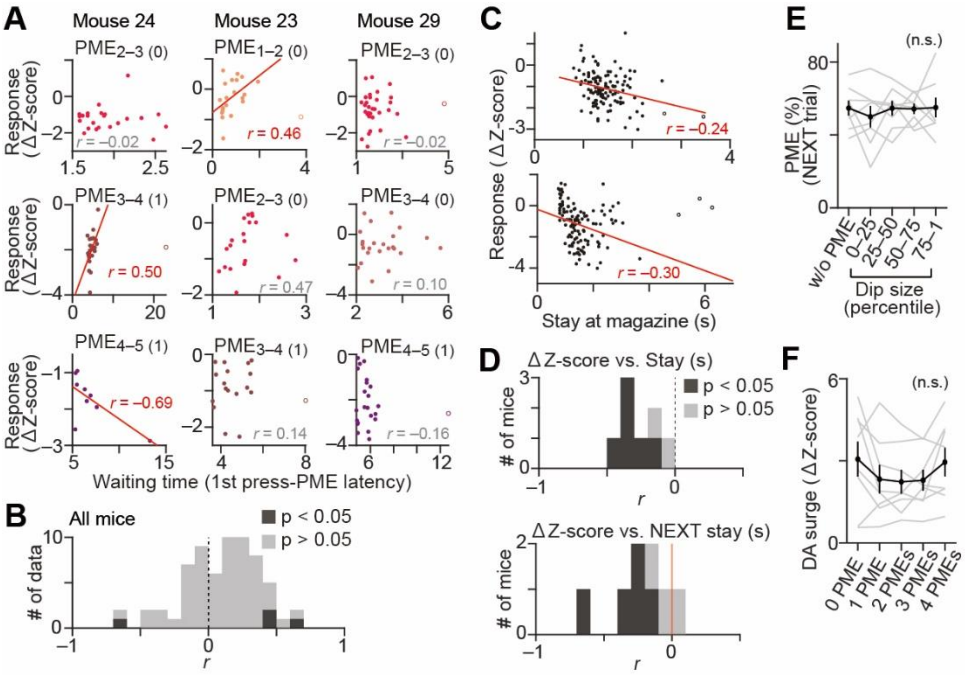

Figure 4—figure supplement 4. DA dip amplitude graduality upon PMEs is not explained by the elapsed time since the mice's task engagement.

(A) Relationship between the latency of a PME from a task engagement (a first press in a given trial) in three representative mice. To exclude the possibility that the DA dip amplitude is correlated with time rather than the number of lever presses, we classified the PME data based on the PME groups ($PME_{1–2}$/$PME_{2–3}$/$PME_{3–4}$/$PME_{4–5}$) and the number of previous PMEs in that trial (number enclosed in parentheses). Three representative data groups from each mouse are shown. Data points outside the mean ± 3 STD were excluded from the correlation analysis (open circles). The Pearson correlation coefficient is shown in each panel. Panel with red line, significant positive or negative correlations. Panel without a line, no significant correlation.

(B) Group data related to ***Fig. S4-4A*** showing the Pearson correlation coefficients (65 data groups out of 80 groups, with 7 or more samples each) for all data groups of all mice tested. No data groups except one showed a significant negative correlation between the DA dip size and the time elapsed, which excludes the possibility that the waiting time explains the variability in the DA dip amplitude across the PME groups. Shade in dark gray, data group with $p < 0.05$. Shade in light gray, data group with $p > 0.05$.

(C) Representative data from two mice showing that a longer stay at the magazine upon a PME coincided with a larger (more negative) DA response magnitude (Pearson correlation coefficient, Mouse 9, n = 127 events, $r = –0.24$, $p = 8.2 \times 10^{–3}$; Mouse 29, n = 118 events, $r = –0.30$, $p = 1.3 \times 10^{–3}$). Data outside the mean ± 3 STD were excluded from the correlation analysis (filled circles).

(D) Top: Group data related to ***Fig. S4-4C*** showing the Pearson correlation coefficients (8 sessions from 8 mice). Six out of eight mice showed significant negative correlations ($p < 0.05$, shaded in dark gray), and all mice showed negative correlation coefficients, excluding the possibility that PMEs provided aversive experiences so that the mice quit the magazine entry earlier if the DA responses were more negative. Bottom: Group data same as the top panel but showing the Pearson correlation coefficients between the DA response amplitude upon a PME and the duration of stay at the magazine upon a *subsequent* PME in the same trial. Five out of

eight mice showed significant negative correlations ($p < 0.05$, shaded in dark gray), and seven mice showed negative correlation coefficients, indicating that the duration of stay at the magazine following a bigger DA dip tended to be longer as well.

(E) Relationship between the DA dip amplitude upon a PME in a trial and the PME percentage in the next trial. The bigger DA dips did not decrease the PME percentage with the corresponding press number, suggesting that the effects of DA dips on the subsequent behavior cannot be explained by simple framing based on classical reinforcement learning theory (n = 5 data groups with 8 sessions each from 8 mice, one-way repeated-measures ANOVA, $F_{(4,28)} = 0.39$, $p = 0.82$). The plotted results correspond to the average values among the inter-press interval groups per dip size percentile in the selected FR5 session. Gray line, individual mouse. Black line, mean ± SEM.

(F) The DA surge amplitude was not sensitive to how efficiently the mice completed a FR5 trial, suggesting that behavior adjustment during the FR5 sessions was not realized simply via the refinement of performance by reinforcing a good performance with fewer PMEs (n = 5 data groups with 8 sessions each from 8 mice, one-way repeated-measures ANOVA, $F_{(4,28)} = 1.60$, $p = 0.20$). Gray line, individual mouse. Black line, mean ± SEM.

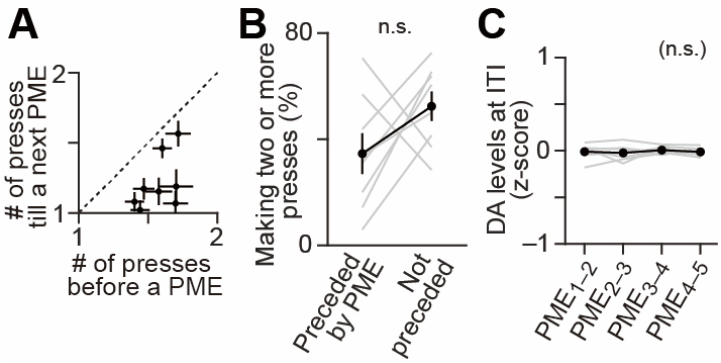

Figure 4—figure supplement 5. Supplementary analyses of behavior and DA dynamics upon PMEs.

(A) Comparison between the number of lever presses before a PME and the number of lever presses after that PME within a trial did not suggest regret upon a PME. Previous studies distinguished disappointment and regret following a recognition that a reality was worse than expected[1,2]. Regret arises if the present unfavorable situation is attributable to its own bad action selection in the past and if an alternative action selection would make the present situation better. If mice in the current study regretted not making enough number of lever-presses before a PME, mice would press the lever more times in a row after the PME[3]. Therefore, we compared the number of consecutive lever-presses before and after a PME. We found that the number of consecutive lever-presses after a PME was significantly smaller than the number of consecutive lever-presses before the PME (n = 8 sessions from 8 mice, one-sample t-test, 0.36 ± 0.06 fewer lever presses after a PME than before the PME, $t_{(7)} = -6.05$, $p = 5.15 \times 10^{-4}$), implying that mice experienced disappointment rather than regret in our study.

(B) Comparison of the probability of multiple lever presses with and without a preceding PME did not suggest regret upon a PME. If mice regretted upon a PME, the mice would make more consecutive lever presses after the PME[2]. Thus, we compared lever presses with the same press number in different trials such as the second, third, or fourth presses with or without a PME to exclude the effects of rising expectations with increasing proximity to task completion. We found that the presence of a preceding PME did not change the percentage of multiple lever presses (n = 8 sessions each from 8 mice, preceded by PME, 34.5 ± 7.7% of the total case; not preceded by PME, 52.4 ± 5.5%; paired t-test, $t_{(7)} = -1.60$, p = 0.16), suggesting that a mouse did not regret upon a PME. These results implied that mice were disappointed rather than regretful upon a PME. Gray line, individual mouse. Black line, mean ± SEM.

(C) DA levels during the ITI periods (–15 to 0 s to sound cue onsets) did not show significant variances among PME group numbers (n = 8 sessions each from 8 mice, one-way repeated-measures ANOVA, $F_{(3,21)} = 0.4$, p = 0.76). Gray line, individual mouse. Black line, mean ± SEM.

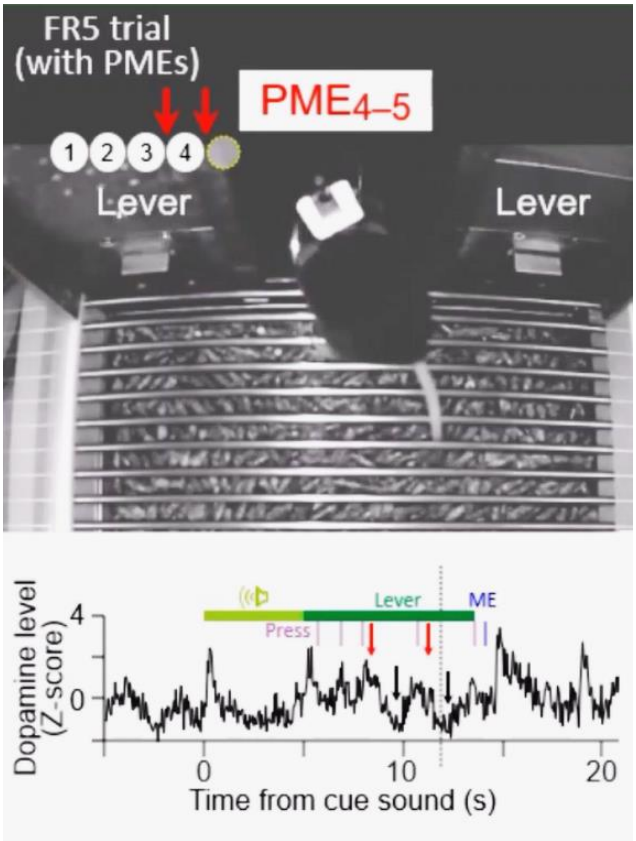

Figure 4—video 1. Phasic dopamine decrease upon a premature head entry to the magazine during the FR5 task.

(Top) Representative behavior of a mouse engaging in the FR5 task. After the 5-s noise sound cue, reinforced (left) and non-reinforced (right) levers were presented. The mouse was rewarded at the magazine (white dashed line) after pressing the reinforced lever five times per trial. Of note, the mouse occasionally made a premature magazine entry (PME) between two consecutive lever presses. $PME_{4\text{-}5}$, a PME between the fourth and the fifth lever presses.

(Bottom) Simultaneous measurement of extracellular DA fluctuations. Vertical dotted line indicates behavior timing. DA level decreased transiently (black arrows just above the DA trace) following a PME (red arrows). We found that amplitudes of a DA dip upon a PME within a session increased with increasing proximity to the task completion, e.g., a DA dip upon $PME_{4\text{-}5}$ was, on average, larger than that upon $PME_{3\text{-}4}$.

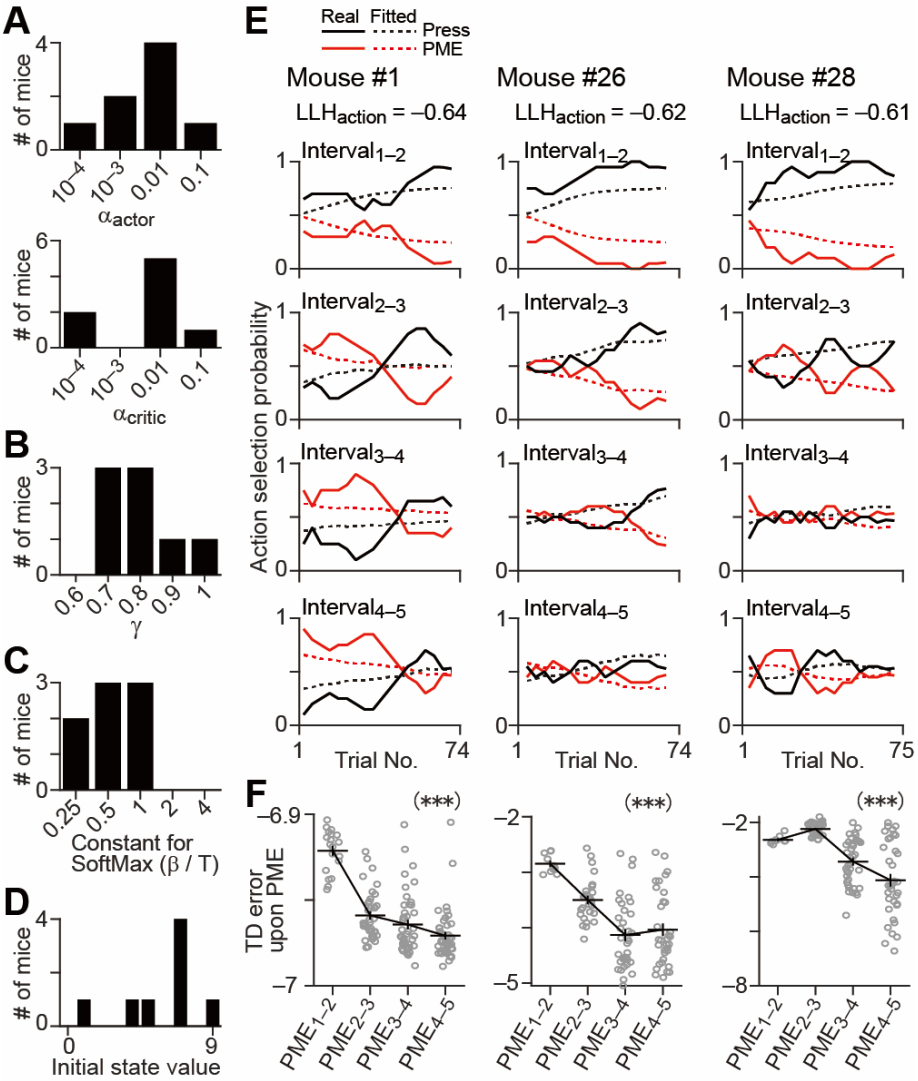

Figure 5—figure supplement 1. RL model simulation.

(A) Group data showing the learning rates (α_{actor} [left] and α_{critic} [right]) used to calculate the TD errors for each step in a FR5 trial in the best-fitted model of the RL algorithm.

(B) Related to ***Fig. S5A*** but for γ, the temporal discount factor used in the calculation of TD errors.

(C) Related to ***Fig. S5A*** but for β/T, a constant in the SoftMax function to compute the action selection probability.

(D) Related to ***Fig. S5A*** but for the initial state values assigned for the grids in the lever zone at the beginning of the FR5 schedule.

(E) Related to ***Fig. 5D*** but showing the observed (solid lines) and simulated (dashed lines) action selection probability changes across the completed trials in the best-fitted model for other representative mice.

(F) Same as ***Fig. 5H***, showing statistical analyses in the TD error amplitude in the best-fitted model for other mice. The mice's number corresponds to the above plots (***Fig. S5E***). (Mouse 1, one-way ANOVA, $F_{(3,146)} = 50.2$, $p = 3.1 \times 10^{-22}$; Mouse 26, one-way ANOVA, $F_{(3,105)} = 18.3$, $p = 1.4 \times 10^{-9}$; Mouse 28, one-way ANOVA, $F_{(3,117)} = 26.9$, $p = 3.2 \times 10^{-13}$). Each open circle represents an individual event. Black line, mean ± SEM. *** $p < 10^{-8}$.

Supplemental References

1 Zeelenberg, M., Dijk, W. W. v., S.R.Manstead, A. & Pligt, J. d. The Experience of Regret and Disappointment. *Cognition Emot* **12**, 221-230, doi:10.1080/026999398379727 (2010).

2 Steiner, A. P. & Redish, D. A. Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. *Nat. Neurosci.* **17**, 995-1002, doi:10.1038/nn.3740 (2014).

3 Dijk, W. W. v., Pligt, J. v. d. & Zeelenberg, M. Effort Invested in Vain: The Impact of Effort on the Intensity of Disappointment and Regret. *Motiv. Emotion* **23**, 203-220, doi:10.1023/a:1021315314979 (1999).