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. We propose that the degree of disappointment during making an effort for 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, experimental investigations are limited regarding the neuronal representation of disappointment, a worse-than-expected outcome inducing a negative RPE. 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 by previous studies 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) is a standard method for the direct detection of extracellular DA dynamics; however, its sensitivity and stability for DA measurement are comprised of intrinsic noises derived from an oxidation-reduction reaction (***Heien et al., 2005; Phillips et al., 2003; Robinson et al., 2003; Rodeberg et al., 2017***). Due to these difficulties, there is still an ongoing debate as to whether negative RPEs are represented by the tonic or phasic modulation of DA release in the ventral striatum (***Hamid et al., 2016***). Recently, a genetically encoded optical DA sensor, $GRAB_{DA2m}$ (G protein-coupled receptor-activation-based DA), (***Sun et al., 2018; Sun et al., 2020***) was reported, with 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).

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. During the FR5 task, the mice had to press a lever five times to obtain a food reward (***Yoshida et al., 2019***). The advantage of this simple operant task is that it brought various degrees of disappointment to the mice without modulating the reward size or operant condition, either of which might 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. A 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, these findings 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*** and ***1B***). 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 versus FR1-Last session, n = 4 sessions with 4 mice each, Tukey-Kramer post-hoc analysis, $p = 1.6 \times 10^{-2}$). 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) and 2) the number of incomplete trials per session (***Fig. 1D***, right; FR1-First versus FR1-Last session, n = 4 sessions with 4 mice each, Tukey-Kramer post-hoc analysis, first press latency, $p = 4.4 \times 10^{-3}$; number of incomplete trials, $p = 1.5 \times 10^{-2}$).

After the establishment of the lever press–reward delivery relationship by the seven FR1 sessions, we increased the lever press number required for the food reward delivery from one to five. This schedule transition from FR1 to FR5 temporarily decreased the number of completed trials, but this was recovered to a value comparable to that of FR1-Last by the end of the FR5 schedule, implying behavioral adjustment in the mice during the six consecutive FR5 sessions (***Fig. 1D***; n = 4 sessions with 4 mice each, Tukey-Kramer post-hoc analysis, FR1-Last versus FR5-First sessions, $p = 1.6 \times 10^{-2}$; FR1-Last versus FR5-Last sessions, p = 0.21). 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, and did not change by the schedule transition nor throughout the FR5 sessions (***Fig. 1D***, center and right, ***Fig. S1A***; n = 4 sessions with 4 mice each, Tukey-Kramer post-hoc analysis, FR1-Last versus FR5-First sessions, p = 0.85 [first press latency], p > 0.99 [number of incomplete trials]; FR1-Last versus FR5-Last sessions, p = 0.77 [first press latency], p > 0.99 [number of incomplete trials]). 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***, n = 6 sessions with 4 mice each, one-way repeated measures ANOVA: $F_{(5,15)} = 7.78$, $p = 8.7 \times 10^{-4}$). Note that the percentage of incomplete trials per session were negligible after the mice had established the lever press–reward delivery relationship during the FR1 schedule (0.5 ± 0.3% and 0.7 ± 0.4% of all trials in the FR1-Last and FR5-First sessions, respectively, n = 4 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, and 2) the mice adjusted themselves to the FR5 sessions to efficiently attain beneficial outcomes by shortening their task completion duration.

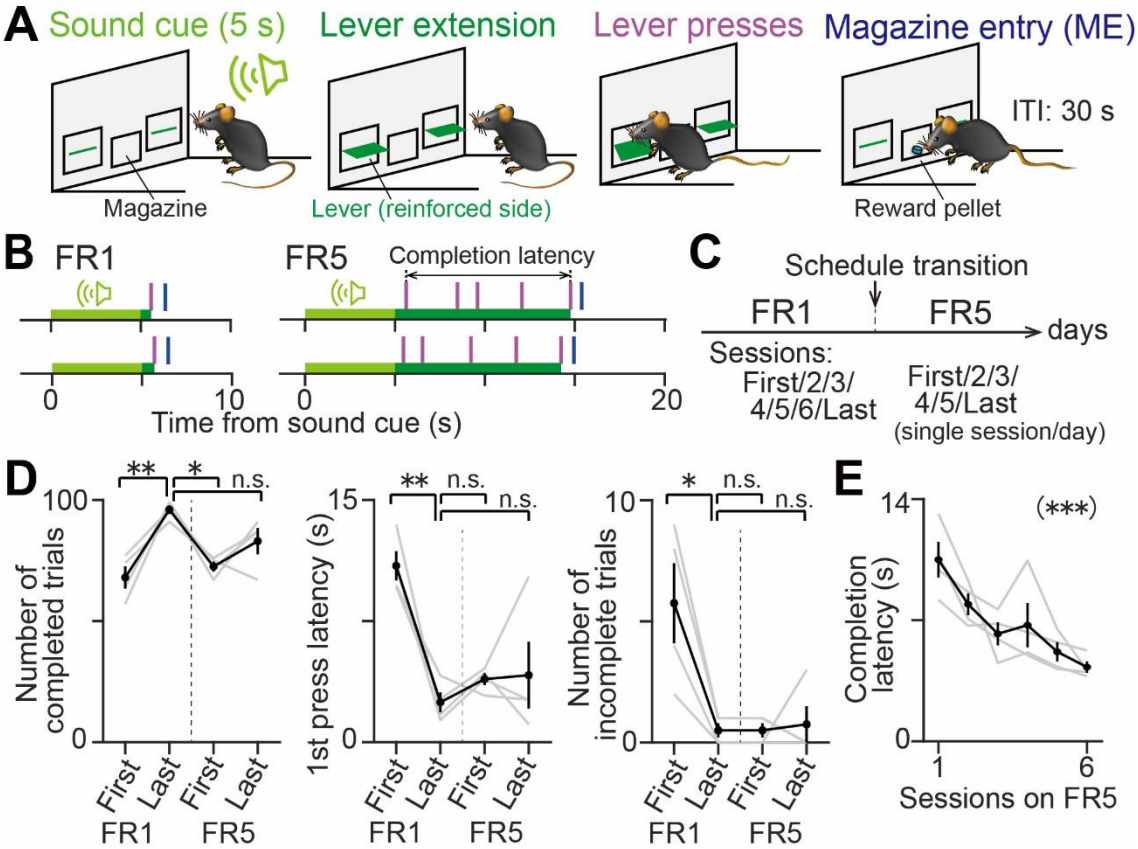

Figure 1. Behavior adjustment across FR5 sessions manifested in completion latency.

(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 4 mice each, Tukey-Kramer post-hoc analysis, FR1-First versus FR1-Last, $p = 5.5 \times 10^{-3}$ [number of completed trials], $p = 4.4 \times 10^{-3}$ [first press latency], $p = 1.5 \times 10^{-2}$ [number of incomplete trials]). While the

number of completed trials transiently decreased with a schedule transition from FR1 to FR5 (vertical dotted lines in D), all three parameters failed to detect behavior adjustment during the FR5 sessions (FR1-Last versus FR5-First, $p = 1.6 \times 10^{-2}$ [number of completed trials], $p = 0.85$ [first press latency], $p > 0.99$ [number of incomplete trials]; FR1-Last versus FR5-Last, $p = 0.21$ [number of completed trials], $p = 0.77$ [first press latency], $p > 0.99$ [number of incomplete trials]). * $p < 0.05$. ** $p < 10^{-2}$.

(E) Completion latency, the time from the first to last lever press (***Fig. 2B***, right), detected behavior adjustment during FR5 sessions (n = 6 sessions with 4 mice each, one-way repeated measures ANOVA, $F_{(5,15)} = 7.78$, $p = 8.7 \times 10^{-4}$). Gray line, individual mouse. Black line, mean ± SEM. *** $p < 10^{-3}$.

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, which we called premature magazine entry (PME; ***Fig. 2A***). One PME occasionally occurred during an inter-press interval between two consecutive lever presses (***Fig. 2B***). Hence, the inter-press intervals were categorized 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). We observed that the inter-press intervals with a PME were significantly longer than those without a PME in all inter-press interval groups (***Fig. 2C*** and ***2D***; n = 15–57 events, Kolmogorov–Smirnov test, all $D_{max} > 0.98$, all $p < 1.1 \times 10^{-11}$), indicating that a PME is a suboptimal behavior that retards task completion. The length of stay at the food magazine following a PME was similar in the four inter-press interval groups (***Fig. S2B***), indicating similar PME motion regardless of groups. We confirmed that variations in the inter-press intervals were explained mostly by the presence or absence of a PME rather than by the session number or inter-press interval group (***Fig. 2E***; n = 3 factors with 4 mice each, one-way repeated measures ANOVA, $F_{(2,6)} = 93.82$, $p = 3.0 \times 10^{-5}$; Tukey-Kramer post-hoc analysis, with/without PME versus session number, $p = 5.9 \times 10^{-5}$; with/without PME versus inter-press interval group, $p = 4.9 \times 10^{-5}$; session number versus inter-press interval group, $p = 0.93$). The length of stay at the food magazine after a ME was significantly longer than that after a PME, which reflects the consumptive behavior of the mice at the magazine after a ME (***Fig. S2A***).

We examined PMEs in each inter-press interval group to evaluate the mice's behavior adaptation to the FR5 schedule in a representative session (***Fig. 2F***). We observed that 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. The percentage of inter-trial intervals with the mice's PMEs in all the completed trials (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; n = 24 data points, 4 inter-press interval groups × 6 sessions, two-way factorial ANOVA, inter-press interval group, $F_{(3,23)} = 18.42$, $p = 2.7 \times 10^{-5}$; session, $F_{(5,23)} = 26.13$, $p = 6.7 \times 10^{-7}$), which supports the idea that the PME percentage was a sensitive measure for behavior adjustment during the FR5 sessions.

To analyze the PME group data, we used another measure, PME bias, which 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; the more a PME occurred in the first or last half of the groups, the closer the PME bias was to –1 or 1, respectively. The PME bias was 0 if a PME occurred in equal numbers in the first and second half of groups. Using data from all the mice (n = 6 sessions with 4 mice each), we plotted the PME bias versus the overall PME percentage, which is the mean of the PME percentage in all inter-press interval groups in a session with the FR5 schedule (***Fig. 2H***). We found 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***; n = 6 sessions with 4 mice each, one-way repeated measures ANOVA, $F_{(5,15)} = 12.00$, $p = 8.4 \times 10^{-5}$), indicating the increasing bias throughout the FR5 sessions (i.e., PMEs occurred more in the last half of the inter-press interval groups when compared to the first half). The overall PME percentage decreased gradually throughout the FR5 sessions (***Fig. 2H***, y-axis; n = 6 sessions with 4 mice each, one-way repeated measures ANOVA, $F_{(5,15)} = 19.53$, $p = 4.4 \times 10^{-6}$), which is in line with the observation that the existence of a PME determined the length of an inter-press interval (***Fig. 2E***) and that the completion latency decreased throughout the FR5 sessions (***Fig. 1E***).

Because a PME is a reward-seeking behavior, the PME percentage can be regarded as a behavioral measure for reward expectation. We quantified the degree of expectation in each inter-press interval group during the FR5 sessions 1 and 2 by normalizing the PME percentage (***Fig. 2I***). The behavior based-reward expectations increased nearer to task completion (n = 4 interval groups with 8 sessions each from 4 mice, FR5-First and FR5-2 sessions from each mouse, one-way repeated measures ANOVA, $F_{(3,21)} = 40.84$, $p = 6.1 \times 10^{-9}$).

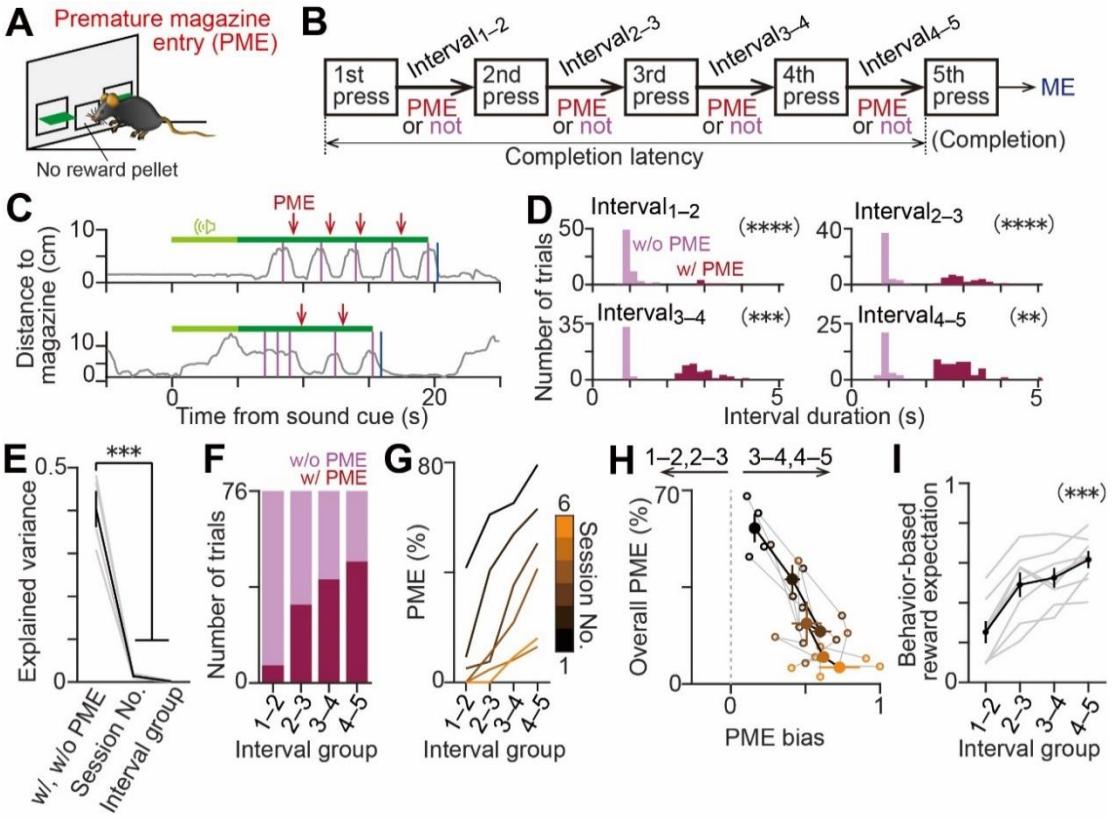

Figure 2. Rising expectations nearer task completion can be quantified by the percentage of a premature magazine entry.

(A) The mice occasionally checked the food magazine without a reward before completing the FR5 task. This behavior was called premature magazine entry (PME).

(B) Scheme of a FR5 lever press task. An inter-press interval, which is a period between two consecutive lever presses, was distinguished by the order of lever presses to be, such as $\text{Interval}_{1\text{–}2}$, which 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, grouped into $\text{Interval}_{1\text{–}2}$, $\text{Interval}_{2\text{–}3}$, $\text{Interval}_{3\text{–}4}$, or $\text{Interval}_{4\text{–}5}$, with (dark red) or without (pale red) PME in a session of mouse 9. Distribution of inter-press interval durations with a PME (dark red) was significantly different from that without a PME (pale red) ($\text{Interval}_{1\text{–}2}$, n = 30 and 42 events with and without PME, respectively, Kolmogorov-Smirnov test, $p = 6.1 \times 10^{-16}$; $\text{Interval}_{2\text{–}3}$, n = 44 and 28 events, $p = 2.3 \times 10^{-16}$; $\text{Interval}_{3\text{–}4}$, n = 47 and 25 events, $p = 1.2 \times 10^{-15}$; $\text{Interval}_{4\text{–}5}$, 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 ($\text{Interval}_{1\text{–}2/2\text{–}3/3\text{–}4/4\text{–}5}$) (n = 3 factors with 4 mice each, Tukey-Kramer post-hoc analysis, PME versus session number, $p = 5.9 \times 10^{-5}$; PME versus inter-press interval group number, $p = 4.9 \times 10^{-5}$).

Gray line, individual mouse. Black line, mean ± SEM. *** $p < 10^{-4}$.

(F) Representative stacked bar chart of mouse 9 during the FR5-2 session. The number of trials with (dark red) or without (pale red) a PME increased and decreased respectively with the interval number increase; the mouse tended to exhibit more PMEs as the FR5 task completion neared (e.g., the number of trials with a PME during $\text{Interval}_{4–5}$ was larger than that during $\text{Interval}_{1–2}$).

(G) Representative session-by-session plot of the PME percentage versus the interval group 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 $\text{Interval}_{4–5}$ was larger than that during $\text{Interval}_{1–2}$ (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 showing the trend that an increase in the session number from the first (black) to the last (brown) session during the FR5 schedule is accompanied by an increase in PME bias (n = 6 sessions with 4 mice each, one-way repeated measures ANOVA, $F_{(5,15)} = 12.00$, $p = 8.4 \times 10^{-5}$) and a decrease in the overall PME percentage ($F_{(5,15)} = 19.53$, $p = 4.4 \times 10^{-6}$). Open circle, individual session (session numbers ere color-coded as in G). Filled circle, mean ± SEM within a session. Gray line, individual mouse. Black line, mean of all mice.

(I) Reward expectations (normalized PME probabilities) of mice during a session on a FR5 schedule increased as the FR5 task neared completion (n = 4 interval groups with 8 sessions each from 4 mice, FR5-First and FR5-2 sessions from each mouse, one-way repeated measures ANOVA, $F_{(3,21)} = 40.84$, $p = 6.1 \times 10^{-9}$). Gray line, individual session. Black line, mean ± SEM. *** $p < 10^{-8}$.

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. Specifically, we used a viral approach to express a genetically encoded optical dopamine sensor—$\text{GRAB}_{\text{DA2m}}$—in the ventrolateral striatum (VLS) (***Fig. 3A***). Additionally, 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 $\text{GRAB}_{\text{DA2m}}$ 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) clearly captured both phasic and tonic DA release over a period of sub-seconds to tens of seconds (***Fig. 3C***). The fluorescent intensity was stable for an hour with little photobleaching effects, indicating stable DA measurements for each whole session (***Fig. S3A*** and ***S3B***). Previous studies with electrophysiological, electrochemical, or optical measurements have reported the activity of midbrain DA neurons representing reward prediction errors (RPEs) (***Amo et al., 2020; Schultz et al., 1997; Starkweather and Uchida, 2021***). Our fiber photometry system is advantageous

in that it enables direct DA observation at a target site of DA neurons with a higher DA specificity and fewer motion artifacts than previous methods.

We observed phasic increases in the extracellular DA levels at the onsets of the sound cue, lever extension, and ME in the FR1 operant task (***Fig. 3D***). The DA increase that accompanied the lever press (magenta line) was not evident in this experiment, which is in accordance with a previous report showing that dopaminergic axons projecting to the ventral striatum were less modulated by mice's motion than those in the dorsal striatum (***Howe and Dombeck, 2016***) (***Fig. 3D***). We next analyzed the modulation of the DA dynamics through the operant task sessions and the schedule transition from FR1 to FR5. Representative peri-event DA signal dynamics from four sessions (FR1-First, FR1-Last, FR5-First, and FR5-Last; see ***Fig. 1D–F***) are plotted in Fig. 3E, and group data in terms of the DA response amplitudes is summarized in Fig. 3F to 3H. There was a significant increase in the DA response to the sound cue from the FR1-First to FR1-Last sessions, indicating an establishment of an association between the lever press and food reward throughout the sessions (***Fig. 3E***, left panel), whereas no significant change was observed between the FR1-Last and FR5-First sessions (***Fig. 3F***, n = 4 sessions with 4 mice each, one-way repeated measures ANOVA, $F_{(3,9)} = 9.71$, $p = 3.5 \times 10^{-3}$; Tukey-Kramer post-hoc analysis, FR1-First vs FR1-Last, $p = 6.4 \times 10^{-3}$; FR1-Last vs FR5-First, p = 1.00). We also detected an immediate elevation of DA levels at the onset of the lever extension (***Fig. 3E***, left, lever extension onset at 5 s), the response amplitudes of which were similar among sessions (i.e., the amplitudes of the DA response to the lever extension in the FR1-First or FR5-First sessions were similar to those under the well-trained condition, the FR1-Last session (***Fig. 3G***, n = 4 sessions with 4 mice each, one-way repeated measures ANOVA, $F_{(3,9)} = 3.2$, $p = 7.6 \times 10^{-2}$).

These DA dynamics in our study were mostly in line with those reported in previous studies of Pavlovian conditioning (***Amo et al., 2020***) or operant conditioning (***Schultz et al., 1997***): 1) at the beginning of the conditioned stimulus (CS)–unconditioned stimulus (US) association (***Fig. 3E*** to ***3H***, FR1-First session in our study), there is a large DA response to the US (the ME in our study, ***Fig. 3E***, right) and an almost negligible response to the CS (the sound cue and lever extension in our study; ***Fig. 3E***, left); 2) once the CS–US association is established (***Fig. 3E***, FR1-Last session in our study), the DA activity shows no obvious responses to the US but clear responses to the CS; 3) upon a schedule transition that violates the CS–US association, patterns of DA response to the US and CS become similar to those during initial periods of the CS–US association (i.e., the response to the US becomes larger and that to the CS diminishes). A minor difference between the previous reports and our current study is that the DA response to the CS (the sound cue or lever presentation) did not diminish and remained high even after the schedule transition from FR1 to FR5 in the current study (***Fig. 3F*** and ***3G***, FR1-Last versus FR5-First), suggesting that the schedule transition did not diminish the association between the cue and reward (represented as PREs in the DA

activity) and explaining why the mice initiated operant responses with a comparable degree of motivation even after the schedule transition (***Fig. 1D***, center and right).

We found that the schedule transition from FR1 to FR5 only modulated the DA response upon the ME (***Fig. 3E***, right, black versus light purple lines). The amplitudes of the DA response upon the ME were comparable between the FR1-First and -Last sessions and increased significantly from the FR1-Last to the FR5-First sessions (***Fig. 3H***; n = 4 sessions with 4 mice each, one-way repeated measures ANOVA, $F_{(3,9)} = 4.3$, $p = 3.8 \times 10^{-2}$; Tukey-Kramer post-hoc analysis, FR1-Last versus FR5-First, $p = 4.2 \times 10^{-2}$; FR1-Last versus FR5-Last, p = 0.94). The fact that only the DA response to the US (magazine entry) was modulated by the schedule transition (***Fig. 3H***) and that the DA response to the CS (sound cue and lever extension) remained stable (***Fig. 3F*** and ***3G***) suggested the importance of the DA signal modulation upon a ME or PME in the mice's behavior adjustment during the FR5 schedule.

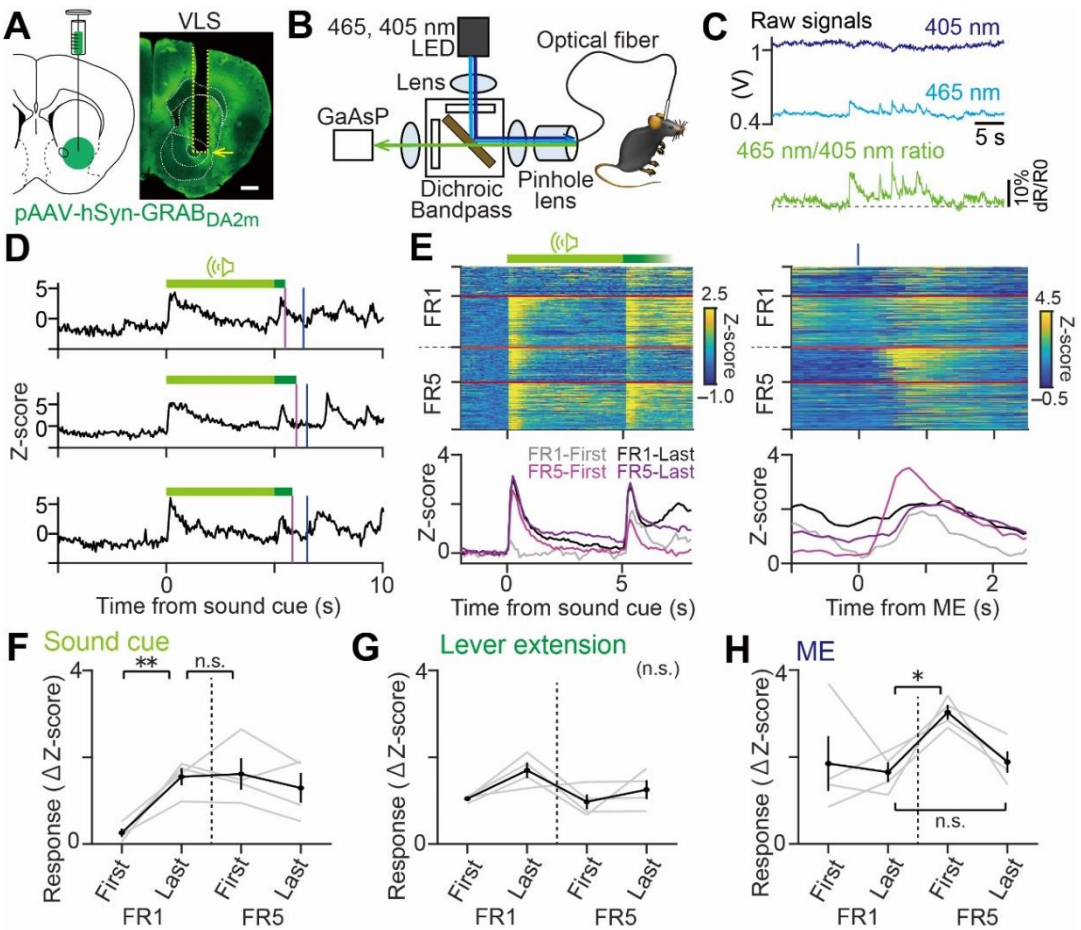

Figure 3. Optical measurement of extracellular DA dynamics in the ventrolateral striatum during a fixed ratio lever press task.

(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 arrowhead, fiber tip. Scale bar, 500 mm.

(B) Schematic of 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) or 405 nm (dark blue line) light. Corresponding ratio values (green; 465 nm/405 nm; shown in a normalized form of dR/R0) indicate extracellular DA dynamics. Gray dashed line, the ratio baseline (R0).

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

(E) Top: Representative heatmaps demonstrating $GRAB_{DA2m}$ signal fluctuations during sound cue periods (left; light green bar) followed by lever extension onsets (left; dark green bar) or ME timings (right; vertical blue tick) 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 gray dashed line along the left axis indicates the schedule transition timing from FR1 to FR5. Bottom: Averaged $GRAB_{DA2m}$ signal traces for respective sessions on the heatmap.

(F-H) $GRAB_{DA2m}$ signal dynamics at an ME (H) captured the schedule transition from the FR1 to FR5 schedule (n = 4 sessions with 4 mice each, Tukey-Kramer post-hoc analysis: FR1-Last versus FR-First at ME, $p = 4.2 \times 10^{-2}$). Contrarily, the DA surge amplitude at the sound cue onset (F) or lever extension (G) did not change throughout the schedule transition (n = 4 sessions with 4 mice each, Tukey-Kramer post-hoc analysis: FR1-Last versus FR5-First upon the sound cue, $p = 1.00$; one-way repeated measures ANOVA at the lever extension, $F_{(3,9)} = 3.2$, $p = 7.6 \times 10^{-2}$). Gray line, individual mouse. Black line, mean ± SEM. * $p < 0.05$, ** $p < 10^{-2}$.

Results 4: DA Dip Amplitudes Quantify Disappointment

We analyzed the DA dynamics during the PME in detail because our results demonstrated that the PME was related to the reward expectations of the mouse (***Fig. 2I***) as well as that the DA dynamics at the food magazine were important for behavior adjustment during the FR5 schedule (***Fig. 3H***). The PMEs were classified into four groups according to the lever press order (e.g., $PME_{2–3}$ stands for a PME which occurred between the first to third lever presses) (***Fig. 4A***).

We noticed a transient drop in the DA levels (***Fig. 4B***, downward black arrow) following a PME (***Fig. 4B***, red arrow) in representative $GRAB_{DA2m}$ signal fluctuation during two FR5 trials. The mean time course of DA signal dynamics of all trials in a session revealed significant phasic decreases (***Fig. 4C***, bottom, red line) or phasic increases (***Fig. 4C***, bottom, blue line) following the PME (regardless of the PME group) or ME, respectively (***Fig. 4D***; PME_{all}, n = 151 events, z-score = –0.65 ± 0.04,

one-sample t-test with Bonferroni correction, $t_{(150)} = -15.14$, $p = 1.0 \times 10^{-31}$; ME, n = 76 events, z-score = 2.91 ± 0.16, $t_{(75)} = 18.10$, $p = 8.6 \times 10^{-29}$). We refer to these phasic DA decreases following a PME and phasic DA increases after a ME as the DA dip and DA surge, respectively, in view of our observation that such phasic activities consistently appeared in the group data of all mice (***Fig. 4E***; PME_{all}, n = 24 sessions each from 4 mice, 6 sessions from each mouse, z-score = –0.91 ± 0.05, one-sample t-test with Bonferroni correction, $t_{(23)} = -17.26$, $p = 2.32 \times 10^{-14}$; ME, n = 24 sessions each from 4 mice, 6 sessions from each mouse, z-score = 1.20 ± 0.14, $t_{(23)} = 8.42$, $p = 3.5 \times 10^{-8}$). These results led us to postulate that a DA dip after a PME represents disappointment, since the DA dip size reflects the expectation amplitude with a reversed sign, based on the following considerations: 1) the RL theory formulates a reward prediction error (RPE) as the difference in the achieved reward and the expected reward; 2) the DA dynamics reflect the RPE; and 3) the achieved reward after the PME is zero (no reward) in the current study. We did not find evidence that PMEs induced regret to the mice (***Dijk et al., 1999; Steiner and Redish, 2014; Zeelenberg et al., 2010***), indicating that PMEs provoked regular disappointing experiences (***Fig. S4A*** and ***S4B***).

We analyzed the quantitative relationship between the DA dip and disappointment by comparing the DA dip amplitudes after a PME with a representative mouse in each PME group (***Fig. 4A***). We discovered that the amplitude of the DA dip was larger with the PME group number. The largest DA dip was observed in the $PME_{4–5}$ group (***Fig. 4F***, left; n = 18−52 events, one-way factorial ANOVA, $F_{(3,150)} = 3.87$, $p = 1.1 \times 10^{-2}$). This proportional relationship between the DA dip amplitude and PME group disappeared with the shuffling of the PME group numbers (***Fig. 4F***, right; $F_{(3,150)} = 0.73$, p = 0.54), indicating that proximity to the completion of the task affected the depth of the DA dip. We confirmed this relationship between the DA dip depth and PME group number using data from all the mice (***Fig. 4G***). There was a consistent gradual increase in the DA dip magnitude with the PME group (***Fig. 4G***; n = 4 PME groups with 8 sessions each from 4 mice, FR5-First and FR-2 sessions from each mouse, one-way repeated-measures ANOVA, $F_{(3,21)} = 22.19$, $p = 1.0 \times 10^{-6}$).

We examined the DA level fluctuations that elicited the distinct magnitudes of a DA dip further in detail. We thought that the dip size differences were caused by trough-level variations (***Fig. 4I***). However, one previous research proposed that the absolute trough level of DA fluctuation was constant but that modulation of a DA baseline preceding the dip caused apparent dip size differences, emphasizing the role of tonic DA fluctuations in achieving a quantitative representation of a negative RPE (***Fig. 4H***) (***Hamid et al., 2016***). We tested these two models by comparing the baseline and trough levels of a DA dip for each PME group. The DA signal fluctuation was adjusted to an ITI period in each trial (–1 to 0 s relative to the sound cue), since there were no significant DA-level variations during the ITI periods among the PME groups (***Fig. S4D***). The baselines level showed only a weak increasing tendency with the PME group (***Fig. 4J***; n = 4

PME groups with 8 sessions each from 4 mice, FR5-First and FR-2 sessions from each mouse, one-way repeated-measures ANOVA, $F_{(3,21)} = 2.91$, $p = 5.9 \times 10^{-2}$). In contrast, we found a significant decrease in the DA trough levels along with the PME group number (***Fig. 4K***; one-way repeated-measures ANOVA, $F_{(3,21)} = 3.4$, $p = 3.7 \times 10^{-2}$). Thus, our observation supports the latter model in that DA level fluctuations in the trough rather than baseline caused the DA dip variation with increasing proximity to task completion.

We evaluated the relationship between the DA dip size and magnitude of expectation by calculating their correlation, discovering a significant positive correlation (***Fig. 4L***; n = 8 sessions with 4 PME groups each, 4 mice in total, FR5-First and FR-2 sessions from each mouse, Pearson correlation coefficient, $r = 0.54$, $p = 1.3 \times 10^{-3}$), demonstrating that extracellular DA concentration quantifies the magnitude of disappointment.

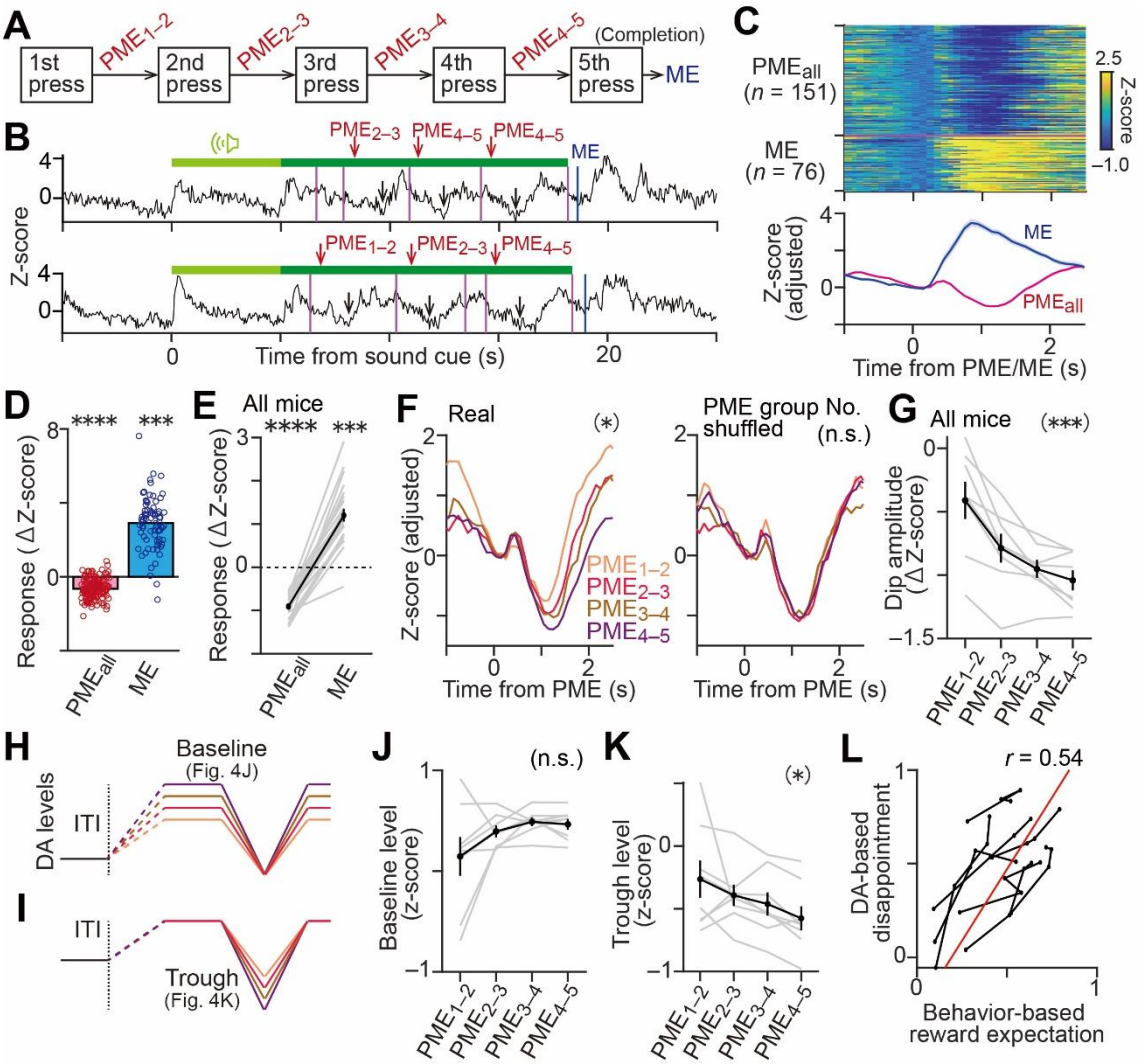

Figure 4. DA dip amplitudes reflect growing disappointment following PMEs as FR5 task completion approaches.

(A) Premature entries (PMEs) were grouped by the order of lever presses (e.g., $PME_{1–2}$,

which was a PME occurring during a period between the first and second lever presses. ME, magazine entry.

(B) Representative $GRAB_{DA2m}$ signal dynamics of mouse 11 during a trial with PMEs on a FR5 schedule. A few seconds after the PMEs (red arrows), a DA dip occurred (black arrows). The DA dip increased as the PME group number increased (e.g., the DA dip following $PME_{4–5}$ was deeper than that following $PME_{1–2}$). Light and dark green bars, sound cue and lever extension periods. Vertical magenta and blue lines, lever press and magazine entry (ME), respectively.

(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} shows the DA dynamics upon the PMEs regardless of the PME group number. Bottom: Mean $GRAB_{DA2m}$ signal traces for PME_{all} and ME on the heatmap with a bin size of 100 ms (shown as mean ± SEM). DA surge and dip following a ME (blue) and PME (red), respectively, are evident.

(D) Representative amplitudes of a DA dip after a PME (red) or that of a DA surge after a ME (blue) of mouse 11 shown in ***Fig. 4C*** (PME_{all}, n = 151 events, z-score = –0.65 ± 0.04, one-sample t-test with Bonferroni correction, $t_{(150)} = –15.14$, $p = 1.0 \times 10^{–31}$; ME, n = 76 events, z-score = 2.91 ± 0.16, $t_{(75)} = 18.10$, $p = 8.6 \times 10^{–29}$). Open circle, individual event. *** $p < 10^{–28}$. **** $p < 10^{–30}$.

(E) Group data of DA fluctuations confirmed a significant DA dip and surge after a PME and ME, respectively (one-sample t-test with Bonferroni correction: PME_{all}, n = 24 sessions each from 4 mice, 6 sessions from each mouse, z-score = –0.91 ± 0.05, $t_{(23)} = –17.26$, $p = 2.32 \times 10^{–14}$; ME, n = 24 sessions each from 4 mice, 6 sessions from each mouse, z-score = 1.20 ± 0.14, $t_{(23)} = 8.42$, $p = 3.5 \times 10^{–8}$). Gray line, individual session. Black line, mean ± SEM. *** $p < 10^{–7}$. **** $p < 10^{–13}$.

(F) Representative $GRAB_{DA2m}$ traces during session 1 of mouse 11, plotted for each PME group. Left: DA dip amplitudes following PMEs 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: This trend disappeared if the PME group numbers were shuffled (one-way ANOVA: Session 1 of mouse 11; real, $F_{(3,150)} = 3.87$, $p = 1.1 \times 10^{–2}$; shuffled, $F_{(3,150)} = 0.73$, $p = 0.54$). * $p < 10^{–2}$.

(G) Group data comparison of DA dip amplitudes among PME groups confirmed deeper DA dips with increasing PME group number (n = 4 PME groups with 8 sessions each from 4 mice, FR5-First and FR-2 sessions from each mouse, one-way repeated-measures ANOVA, $F_{(3,21)} = 22.19$, $p = 1.0 \times 10^{–6}$). Gray line, individual session. Black line, mean ± SEM. *** $p < 10^{–5}$.

(H, I) A baseline model (H) and trough-level model (I), either of which may explain the observation of distinct DA dip amplitudes among the PME groups. In the baseline model, deeper DA dips during larger PME group numbers were achieved with higher DA baselines, while the trough levels of DA were the same among the PME groups (e.g., the baseline DA levels were higher during $PME_{3–4}$ than $PME_{1–2}$). In the trough-level model, the DA trough levels were deeper in correspondence with larger PME group numbers, while the baseline DA levels were the same among all PME group numbers. Line colors indicate PME group number, as in ***Fig. 4F***. For comparison, each $GRAB_{DA2m}$ trace was adjusted to an ITI period.

(J) Similar baseline levels before DA dips among different PME group numbers—not in line with the baseline model (one-way repeated-measures ANOVA, $F_{(3,21)} = 2.91$, p = 5.9×10^{-2}). Gray line, individual session. Black line, mean ± SEM.

(K) Significantly different trough levels of DA dips among different PME group numbers, supporting the trough-level model (one-way repeated-measures ANOVA, $F_{(3,21)} = 3.4$, p = 3.7×10^{-2}). Gray line, individual session. Black line, mean ± SEM. *** $p < 10^{-4}$.

(L) Positive correlation between the magnitude of expectations represented by behavior (***Fig. 2I***) and the magnitude of disappointment represented by DA dips (***Fig. 4F***) (n = 8 sessions with 4 PME groups each, 4 mice in total, FR5-First and FR-2 sessions from each mouse, Pearson correlation coefficient, $r = 0.54$, p = 1.3×10^{-3}). Each dot represents a result from a PME group, such as $PME_{3–4}$, in a single session. Each black line connects data within the same session.

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

We employed a computational modeling of the behavioral data to address the question of whether the DA dip size variation had any functional role in behavior adjustment during the FR5 schedule. We used a RL model with actor-critic methods with a temporal difference (TD) error, because it has been proposed that phasic DA signals function as TD errors, 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***). In the model, a lever press and the PME behaviors of the mice during the FR1 and FR5 schedules were simulated in a two-dimensional grid world environment (***Fig. 5A***). For each lever zone (***Fig. 5A***, white squares in the grid world environment), the agents hold a state value and action values for a rightward step (lever press) and for an upward step (ME or PME). The agents then begin at the starting zone on the far left in the grid world environment (***Fig. 5A***) and chooses an action for a rightward or upward step based on probabilities determined via the SoftMax decision rule using the action values (***Fig. 5B***). The agents then update the state and action values step-by-step using a state value-derived TD error (see the Methods section "Simulation of Behavior and Dopamine Dip").

We compared four mathematical models to determine if they could reproduce fluctuations in 1) the PME percentage (***Fig. 2G*** and ***2H***) and 2) the DA dip amplitude (***Fig. 4G*** and ***4L***) throughout the FR5 sessions. We first analyzed two models: deterministic models with (***Fig. 4D***) and without (***Fig. 4C***) pre-training. The term "deterministic" means that the agents' movement between lever zones is faithful to the agents' choice of action (e.g., if the agents choose an action for a rightward step (a lever press), they move to the next lever zone on the right without skipping some zones). The term "pre-training" means that the agents update the state and action values—which are initially set to 0—during the simulation for a FR1 schedule before the FR5 schedule. We found that the first model, the deterministic model without pre-training, failed to reproduce the animal behavior (***Fig. 2G*** and ***2H***). Specifically, the PME percentage of the model did not increase with the inter-press interval group nor decrease with the FR5

session number in a representative agent (***Fig. 5C***, left; n = 24 data points, 4 interval groups × 6 sessions, two-way factorial ANOVA, interval group, $F_{(3,23)} = 18.73$, $p = 2.5 \times 10^{-5}$, session, $F_{5,23} = 0.08$, $p > 0.99$). In the group data of 4 agents, the PME bias of the model decreased rather than increased with the FR5 session number (***Fig. 5C***, right, x-axis; n = 6 sessions with 4 agents each, one-way repeated measures ANOVA, $F_{(5,18)} = 65.90$, $p = 6.2 \times 10^{-11}$). Therefore, we tried the second model, the deterministic model with pre-training, which reflected the actual animal training in the FR1 sessions before the mice subjected to the FR5 sessions. The pre-training resulted in a higher action value for a lever press than for a PME (***Fig. S5A*** and ***S5C***). Hence, the model with pre-training reproduced the decreasing trend in the overall PME percentage (representative agent, ***Fig. 5D***, left, y-axis; group data, ***Fig. 5D***, right, y-axis; n = 6 sessions with 4 agents each, one-way repeated measures ANOVA, $F_{(5,18)} = 529.07$, $p = 7.2 \times 10^{-19}$). However, the PME bias of the model decreased rather than increased with the FR5 session number (***Fig. 5D***, right, x-axis; $F_{(5,18)} = 30.94$, $p = 3.1 \times 10^{-8}$).

Since the deterministic models did not reproduce the experimentally observed behavior adjustment during the FR5 sessions, we next considered a non-deterministic, state transition model (***Starkweather et al., 2018***). The model assumes that the mice sometimes misunderstand the lever press number (***Fig. 2B***) (e.g., the mice may mistake the second lever press for the fourth lever press). To implement this stochastic error of the mice in the model, we had the movement of the agents fluctuate between the lever zones. Specifically, the agents' action for 1) a rightward step resulted in moving from one to three lever zone(s) to the right, and 2) a PME was followed by the agent's movement from one to three lever zone(s) to the left (***Fig. 5H***). We analyzed this third model, a state transition model without the pre-training on a FR1 schedule (***Fig. 5E***). While the model succeeded in reproducing the increasing trend in the PME bias with the increasing FR5 session number (***Fig. 5E***, right, x-axis; n = 6 sessions with 4 agents each, one-way repeated measures ANOVA, $F_{(5,18)} = 38.40$, $p = 5.5 \times 10^{-9}$), the PME percentage of the model did not decrease with the session number (representative agent, ***Fig. 5E***, left, y-axis; group data, ***Fig. 5E***, right, y-axis; $F_{(5,18)} = 33.93$, $p = 1.5 \times 10^{-8}$). Therefore, we analyzed the fourth model, a state transition model with pre-training on a FR1 schedule (***Fig. 5F***, ***Fig. S5B*** and ***S5D***). We found that this model successfully reproduced the mice's behavioral adjustment during the FR5 sessions. In a representative agent, the PME percentage showed a decreasing trend with the FR5 sessions (***Fig. 5F***, left, x-axis, from black to brown lines) and an increasing trend with the inter-press interval group number (***Fig. 5F***, x-axis; n = 24 data points, 4 interval groups × 6 sessions, two-way factorial ANOVA, interval group, $F_{(3,23)} = 18.42$, $p = 2.7 \times 10^{-5}$; session, $F_{5,23} = 26.13$, $p = 6.7 \times 10^{-7}$). In the group data, the model reproduced the tendency of increasing PME bias and decreasing overall PME percentage with the FR5 session number (***Fig. 5F***, right; n = 6 sessions with 4 agents each, one-way repeated measures ANOVA, overall PME percentage, $F_{(5,18)} = 480.89$, $p = 1.7 \times 10^{-18}$;

PME bias, $F_{(5,18)} = 9.92$, $p = 1.1 \times 10^{-4}$). We confirmed the highest reproducibility of the last model out of the four models by calculating the distance between the trajectories of the actual mice behavior adjustment data and the simulated data on the plane of the PME bias versus overall PME percentage (***Fig. 5G***; n = 4 models with 4 cohorts each, 4 agents in each cohort, one-way factorial ANOVA, $F_{(3,15)} = 2807.30$, $p = 2.4 \times 10^{-17}$; Tukey-Kramer post-hoc analysis, state transition model with pre-training versus the other models, all $p < 10^{-8}$).

We showed that the fourth model reproduced the behavior adjustment during the FR5 sessions. An appropriate fluctuation in the TD error was critical for behavior adjustment in this model. To investigate the functional relevance of DA dips in terms of behavior adjustment, we compared the TD error fluctuation in the model and the DA dip dynamics in the mice. We found that absolute values of both the TD error and DA dip decreased in the same way with the FR5 trials (***Fig. 5I*** and ***5J***), suggesting that the negative reward prediction error decreased as the agents and mice became more experienced with the FR5 sessions. We also found that the absolute value of the TD error after a PME was larger with the PME group on the FR5 schedule (***Fig. 5K***, decreasing lines; n = 4 PME groups with 8 sessions each from 4 agents, first and second FR5 sessions from each agent, one-way repeated-measures ANOVA, $F_{(3,27)} = 37.68$, $p = 8.6 \times 10^{-10}$), which corresponded to the larger DA dip values with increasing PME group number in the actual mice (***Fig. 4F***). Furthermore, the positive correlation between the animal behavior and DA dynamics (***Fig. 4L***) was reproduced with this model (***Fig. 5L***; n = 8 sessions with 4 PME groups each, 4 agents in total, first and second FR5 sessions from each agent, Pearson correlation coefficient, $r = 0.92$, $p = 9.4 \times 10^{-14}$), suggesting that the representation of larger disappointment with a larger negative RPE mediated behavior adjustment during the FR5 sessions.

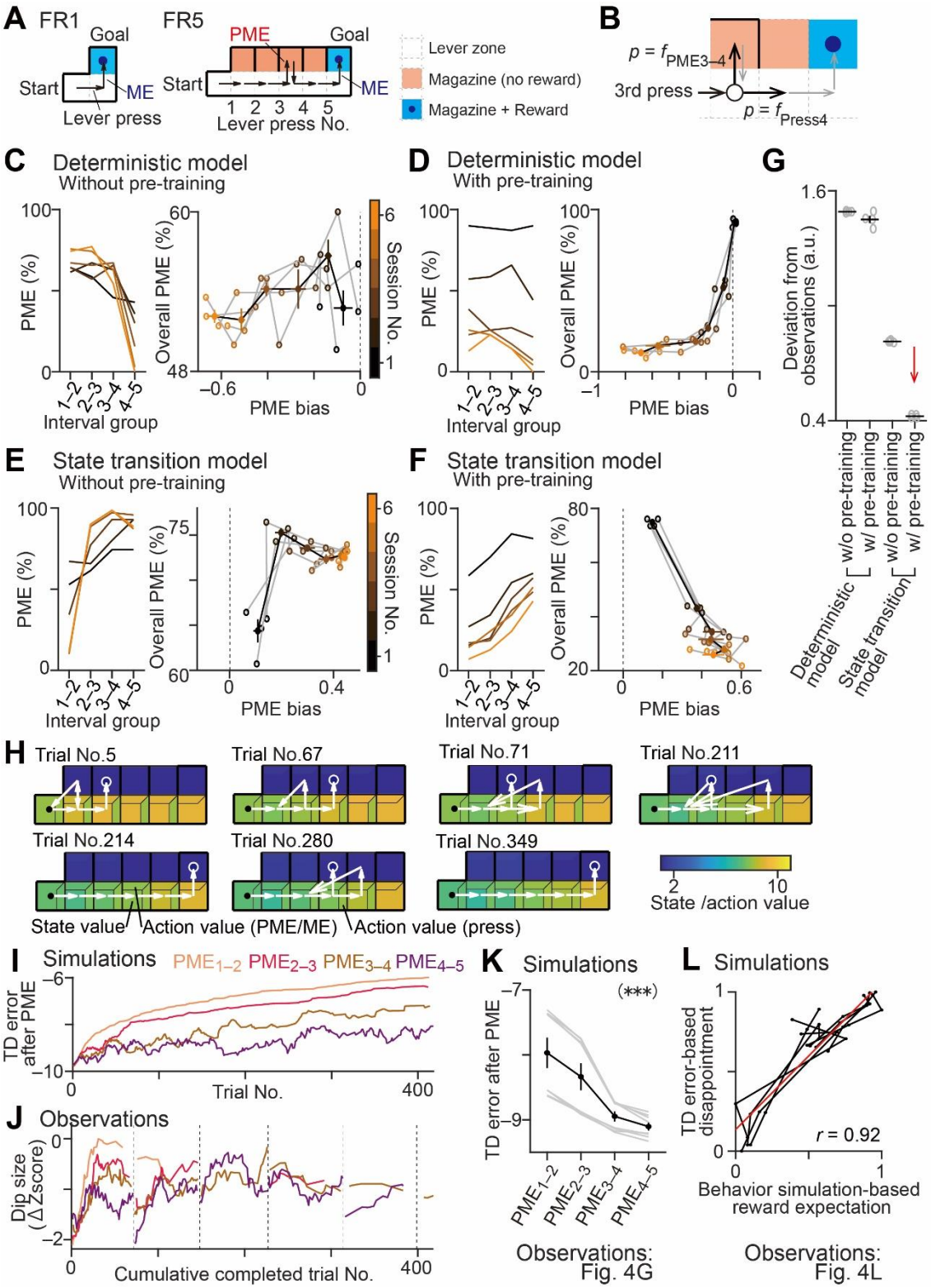

Figure 5. A computational model replicates the increasing trend of behavior-based expectation and DA-based disappointment near FR5 task completion.

(A) Two-dimensional grid world environments for agents which implement an actor-critic temporal-difference RL algorithm to simulate the FR1 (left) or FR5 (right) schedule. The

agents begin at the leftmost square (Start) and move toward the top right square (Goal). Each horizontal arrow indicates the action in a representative trial of an agent: a lever press. Upward arrows pointing to a red or blue square correspond to other actions—a PME or ME, respectively.

(B) Magnified view of the grid world environment for the FR5 schedule. After the third lever press, the agent (open circle) has choices for its next action: $PME_{3–4}$ (upward black arrow) with a percentage of $f_{PME3–4}$ and the fourth lever press (horizontal black arrow stemming from the agent) with a percentage of f_{press4}. After $PME_{3–4}$, the agent returns to where it was following the third lever press (downward grey arrow).

(C-F) Simulation of a behavior on a FR5 schedule using deterministic (C, D) or state transition models (E, F) with parameters that were adjusted (with pre-training: D, F) or not adjusted (without pre-training: C, E) on a FR1 schedule before the FR5 simulation. Left plots: Session-by-session plots of the PME percentage versus the interval group of representative agents on the FR5 schedule. The models with pre-training (D, F) replicated the experimentally observed decreasing trend in the mean of the PME percentage with the increasing session number (early and late sessions indicated as black and brown lines, respectively) (n = 24 data points, 4 interval groups × 6 sessions, two-way factorial ANOVA; deterministic model with pre-training, $F_{(5,23)} = 117.44$, $p = 1.8 \times 10^{-11}$ [session number]; state transition model with pre-training, $F_{(5,23)} = 66.64$, $p = 1.1 \times 10^{-9}$ [session number]). The state transition models (E, F) reproduced the experimentally observed increasing trend in PME percentage within a session with the increasing interval group number (n = 24 data points, 4 interval groups × 6 sessions, two-way factorial ANOVA; state transition model without pre-training, $F_{(3,23)} = 15.79$, $p = 6.6 \times 10^{-5}$ [interval group]; state transition model with pre-training, $F_{(3,23)} = 69.81$, $p = 4.8 \times 10^{-9}$ [interval group]). Experimental results are shown in ***Fig. 2G***. Right plots: Group data plot showing the trends in the overall PME percentage and the PME bias of agents throughout the simulated FR5 sessions. The deterministic model with pre-training (D) and state transition model with pre-training (F) replicated the experimentally observed declining trend in the overall PME percentage (y-axis, early and late sessions indicated by black and brown lines, respectively) (n = 6 sessions with 4 agents each, one-way repeated measures ANOVA, deterministic model with pre-training, $F_{(5,18)} = 529.07$, $p = 7.2 \times 10^{-19}$; state transition model with pre-training, $F_{(5,18)} = 480.89$, $p = 1.7 \times 10^{-18}$). Meanwhile, the state transition models (E, F) recaptured the experimentally observed growing trend in the PME bias (x-axis; n = 6 sessions with 4 agents each, one-way repeated measures ANOVA, state transition model without pre-training, $F_{(5,18)} = 38.40$, $p = 5.5 \times 10^{-9}$; state transition model with pre-training, $F_{(5,18)} = 9.92$, $p = 1.1 \times 10^{-4}$). Augmentation of the PME bias (x-axis) and decrease in the overall PME percentage (y-axis) with the increasing session number (from black to brown circles) were reproduced only by the state transition model with pre-training (F). The corresponding experimental results are shown in ***Fig. 2H***. Gray lines, individual data. Black lines, mean ± SEM.

(G) The state transition model with pre-training (red arrow) best fitted to the experimental data for behavior adjustment. The distance between the simulated data points and experimental data points on the plane of PME bias versus overall PME percentage was compared among models (n = 4 models with 4 cohorts each, 4 agents in each cohort, Tukey-Kramer post-hoc analysis, state transition model with pre-training versus the other three models, $p < 10^{-8}$). Gray circles, individual data. Black lines, mean ± SEM.

(H) Representative trajectories (white arrows) with a state value (a color-coded square in the lever zone) and action values (top and right edges of the lever zone for a PME and a lever press, respectively) obtained with the state transition model with pre-training. Note that the position of the rewarded zone for each trial, indicated by the white open circle, is not fixed in the upper right zone because of the state transition algorithm in the model (F), which demonstrates that the agents do not always correctly track the lever press number during the FR5 trials.

(I, J) Representative trial-by-trial temporal difference (TD) errors of the state transition model with pre-training (I) and trial-by-trial DA dip amplitudes of mouse 11 (J) after PMEs, plotted for each PME group, demonstrating a decreasing trend in DA dip size throughout the trials and deeper DA dips for larger PME group numbers.

(K) TD error amplitudes of the state transition model with pre-training after PMEs were larger with increasing PME group number (n = 4 PME groups with 8 sessions each from 4 agents, first and second FR5 sessions from each agent, one-way repeated-measures ANOVA, $F_{(3,27)} = 37.68$, $p = 8.6 \times 10^{-10}$), which reproduced the experimental observations (***Fig. 4G***). Gray lines, individual data. Black lines, mean ± SEM.

(L) The state transition model with pre-training succeeded in replicating the positive correlation between the magnitude of the reward expectation (***Fig. 2I***) and the magnitude of disappointment (***Fig. 4G***) (n = 8 sessions with 4 PME groups each, 4 agents in total, first and second FR5 sessions from each agent, Pearson correlation coefficient, $r = 0.92$, $p = 9.4 \times 10^{-14}$). Dots represent the data of individual PME groups, and lines connect data within the same session.

Discussion

In the current study, we addressed the neuronal representation of disappointment and found that the DA dip amplitude in the ventral striatum quantitatively represented the magnitude of disappointment. To achieve this result, we used a simple operant task that had the following advantages: 1) the magnitude of expectation and the subsequent disappointment can be modulated without changing the reward size or animal behavior, and 2) event-triggered analysis can be performed for various magnitudes of disappointment. Our sensitive and stable DA imaging system combined with the simple operant task revealed that greater expectations, which were estimated based on the animals' behavior, were accompanied by greater DA dips in the ventral striatum. The functional relevance of such DA dip variations in behavior adjustment was demonstrated by a computational modeling of the task using a RL algorithm that successfully reproduced the dynamics of increasing disappointment and DA dip nearer task completion.

The DAergic coding of RPE has been examined in a series of studies (***Schultz et al., 1997; Tobler et al., 2005***). However, as most studies have focused on DA surge and a positive RPE in the context of economic tasks, there is a limited number of studies that have investigated DA dip and a negative RPE (in macaques (***Nakahara et al., 2004; Satoh et al., 2003; Tanaka et al., 2019***), rats (***Bayer and Glimcher, 2005; Hamid et al., 2016; Hart et al., 2014***), and mice (***Dabney et al., 2020***)). Furthermore, while it is clear

that a positive RPE is quantitatively represented by various amplitudes of DA surge, there is controversy as to whether a negative RPE is also quantitatively represented by a DA dip (***Bayer and Glimcher, 2005; Satoh et al., 2003; Tanaka et al., 2019***) (***Dabney et al., 2020; Hamid et al., 2016; Hart et al., 2014***). There are two hypotheses for realizing variations in a DA dip amplitude. The first hypothesis assumes that tonic DA modulation explains variation in the DA dip amplitude (***Hamid et al., 2016***). For example, a larger DA dip is only apparently caused by a higher DA baseline before the DA dip, while the lowest value the DA concentration hits after the DA dip is constant. The second hypothesis claims that phasic DA modulation explains a DA dip modulation. For instance, variation in the DA dip amplitudes was realized because the lowest value of the DA concentration after the DA dip was actually variable, while the baseline was constant.

Our results showed that variation in the DA dip amplitude was achieved via the modulation of the trough level of the DA while maintaining the DA baseline values prior to the DA dip constant. This result contradicts the previous report (***Hamid et al., 2016***) demonstrating that the modulation of a DA dip amplitude was achieved by changing the DA baseline prior to the DA dip while maintaining a constant trough level of the DA dip. This apparent discrepancy might be explained by distinct task designs for an animal. In the previous report, the baseline DA levels were modulated by several rewards obtained in the last 10 trials of a lottery economic task. Therefore, a longer duration of the multiple trials was required in the previous report to modulate the animals' expectations, which might have changed the tonic DA levels. In the current study, multiple trials were not necessary to modulate the animals' expectations, because we observed a monotonic increase in expectations in a single trial (i.e., animal expectations increased with the inter-lever press interval group in the FR5 schedule. The two mechanisms for DA dip variations observed in the previous study (tonic DA modulation) and the current study (phasic DA modulation) may indicate the presence of distinct processes in regulating extracellular DA levels. Tonic DA modulation might be realized by the local regulation of DA release in the striatum, which is the projection site of DA neurons in the VTA, because the tonic spike modulation of DA neurons in the VTA has not been observed (***Mohebi et al., 2019***). On the other hand, our results demonstrated the phasic activity of DA neurons in the VTA to realize variable DA dip amplitudes (***Dabney et al., 2020***). These two phenomena may suggest multiple mechanisms in the brain for quantifying disappointment.

Our study is, to our knowledge, the first to apply a RL (RL) algorithm to a fixed ratio of a lever press task, which has been widely used to assess efforts of an animal. Previous studies applied the RL algorithm to a behavioral task with paired and similar behaviors for action selection such as pressing left or right levers for a food reward (***Bayer and Glimcher, 2005; Hart et al., 2014; Nakahara et al., 2004***). In the present study, we found that mice showed a behavior adjustment through an action selection between non-

similar behaviors, a lever press and a PME. We found that 1) behavior adjustment for a FR5 schedule accompanied modulation of probability for a lever press and for a PME, 2) fluctuation of DA dynamics was evident at PME, and 3) variation of a DA dip amplitude was important for the behavior adjustment. These results demonstrate the applicability of the RL algorithm to a wider range of behavioral tasks. 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 may give rise to behavioral adjustment.

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. Water was readily available.

Surgery

A total of four 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 posterior 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 pAAV-hSyn-GRAB-DA4.4-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. 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.

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 stimulating 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 (600 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.

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 chloral hydrate 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. The z-score of the extracellular DA signal was calculated based on the signal fluctuations before the initiation of the session each day (600 s). 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 0−0.3 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:

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

where $nInterval_{1–2}$ refers to the number of inter-press intervals grouped into $Interval_{1–2}$; $nInterval_{1–2\ \text{with PME}}$ 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 $Interval_{3\text{–}4}$ or $Interval_{4\text{–}5}$ is larger [smaller] than that during $Interval_{1\text{–}2}$ or $Interval_{2\text{–}3}$). 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 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.

The step back position (PI_{back}) after the k-th continuous press of all N action blocks was calculated as:

$$PI_{pre} = \sum_{k}^{N} D^{N-k+1}\ nPress_k$$

$$PI_{back}\ \in\ Z, PI_{pre} - 1 <\ PI_{back}\ \leq\ PI_{pre}$$

where D corresponds to the devaluation index (0.52), and $nPress_k$ indicates the total number of presses at the k-th action block. The agent was taken back to the position corresponding to the end of the PI_{back}-th press. If PI_{back} was zero, the agent was taken back to the start position.

The non-reinforced lever press was not taken into account, because the percentage of it was at a negligible level (***Fig. 1F***).

If the model was "trained", the value function on the FR5 schedule was mainly computed according to that on the FR1 schedule in the following ways:

1) The action value for the lever press on the FR1 schedule was applied to the values for the first to fourth lever presses on the FR5 schedule;
2) We set the state value and action value (PME) for the grid where the agents visited after each lever press on the FR5 schedule to the same values as the counterparts on the FR1 schedule: the state value and action value (ME) of the grid following the lever press, respectively; and
3) The state values of the PME grids were set to 0.

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 (***Fig. S4A***), one-sample t-tests with Bonferroni corrections for individual data in two or more sample data sets (***Figs. 4D, 4E, S3A***), paired t-tests for two-sample data sets (***Figs. S2A*** and ***S4B***), and Tukey-Kramer post-hoc analyses for three or more sample data sets (***Figs. 1D, 2E, 3F, 3H,*** and ***5G***). We also employed Kolmogorov-Smirnov tests to compare two distributions (***Fig. 2D***). 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, 5K, S1A, S2B, S4D, S5C,*** and ***S5D***), one-way factorial ANOVA (***Figs. 4F*** and ***S4C***), and two-way factorial ANOVA (***Figs. 2G*** and ***5C–5F***). We computed Pearson correlation coefficients to measure linear correlations between the two parameters (***Figs. 4L*** and ***5L***). 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.1).

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. All authors 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) JP20dm0207069 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, K.F., Uchida, N., and Watabe-Uchida, M. (2020). A gradual backward shift of dopamine responses during associative learning. *bioRxiv*, 2020.2010.2004.325324. Digital Object Identifier: 10.1101/2020.10.04.325324, PMID:

Bayer, H.M., and Glimcher, P.W. (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

Covey, D.P., and Cheer, J.F. (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, C.K., Hassabis, D., Munos, R., and 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, W.W.v., Pligt, J.v.d., and 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:

Hamid, A.A., Pettibone, J.R., Mabrouk, O.S., Hetrick, V.L., Schmidt, R., Weele, C.M., Kennedy, R.T., Aragona, B.J., and Berke, J.D. (2016). Mesolimbic dopamine signals the value of work. *Nature Neuroscience* **19**, 117-126. Digital Object Identifier: 10.1038/nn.4173, PMID: 26595651

Hart, A.S., Rutledge, R.B., Glimcher, P.W., and Phillips, P.E.M. (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, A.S., Ariansen, J.L., Cheer, J.F., Phillips, P.E.M., Wassum, K.M., and Wightman, M.R. (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, M.W., and Dombeck, D.A. (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., and 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., and 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:

Mohebi, A., Pettibone, J.R., Hamid, A.A., Wong, J.-M.T., Vinson, L.T., Patriarchi, T., Tian, L., Kennedy, R.T., and Berke, J.D. (2019). Dissociable dopamine dynamics for learning and motivation. *Nature*, 1-6. Digital Object Identifier: 10.1038/s41586-019-1235-y, PMID: 31118513

Moutoussis, M., Bentall, R.P., Williams, J., and 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., and 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, A.d., Mimura, M., Takata, N., et al. (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, J.v., and Morgenstern, O. (1945). Theory of Games and Economic Behavior. *The Journal of Philosophy* **42**, 550-554. Digital Object Identifier: 10.2307/2019327, PMID:

O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H., and Dolan, R.J. (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, P.E.M., Stuber, G.D., Heien, M.L.A.V., Wightman, R.M., and Carelli, R.M. (2003). Subsecond dopamine release promotes cocaine seeking. *Nature*, 614-618. Digital Object Identifier: 10.1038/nature01476, PMID: 12687000

Robinson, D.L., Venton, B.J., Heien, M.L.A.V., and Wightman, R.M. (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, N.T., Sandberg, S.G., Johnson, J.A., Phillips, P.E.M., and Wightman, M.R. (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

Satoh, T., Nakai, S., Sato, T., and Kimura, M. (2003). Correlated Coding of Motivation and Outcome of Decision by Dopamine Neurons. *The Journal of Neuroscience*, 9913-9923. Digital Object Identifier: 10.1523/JNEUROSCI.23-30-09913.2003, PMID: 14586021

Schultz, W., Dayan, P., and Montague, R.P. (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, S.J., and Uchida, N. (2018). The Medial Prefrontal Cortex Shapes Dopamine Reward Prediction Errors under State Uncertainty. *Neuron*. Digital Object Identifier: 10.1016/j.neuron.2018.03.036, PMID: 29656872

Starkweather, C.K., and 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, A.P., and Redish, D.A. (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, S.F., Luo, Y., Li, F., Wang, H., Yamaguchi, T., et al. (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., et al. (2020). New and improved GRAB fluorescent sensors for monitoring dopaminergic activity in vivo. *bioRxiv*. Digital Object Identifier: 10.1101/2020.03.28.013722, PMID:

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

Tanaka, S., O'Doherty, J.P., and Sakagami, M. (2019). The cost of obtaining rewards enhances the reward prediction error signal of midbrain dopamine neurons. *Nature Communications*, 3674. Digital Object Identifier: 10.1038/s41467-019-11334-2, PMID: 31417077

Tobler, P.N., Fiorillo, C.D., and Schultz, W. (2005). Adaptive Coding of Reward Value by Dopamine Neurons. *Science* **307**, 1642-1645. Digital Object Identifier: 10.1126/science.1105370, PMID: 15761155

Yoshida, K., Drew, M.R., Mimura, M., and Tanaka, K.F. (2019). Serotonin-mediated inhibition of ventral hippocampus is required for sustained goal-directed behavior. *Nature Neuroscience*, 1-8. Digital Object Identifier: 10.1038/s41593-019-0376-5, PMID: 30988523

Zeelenberg, M., Dijk, W.W.v., S.R.Manstead, A., and Pligt, J.d. (2010). The Experience of Regret and Disappointment. *Cognition and Emotion* **12**, 221-230. Digital Object Identifier: 10.1080/026999398379727, PMID:

Supplementary Figures

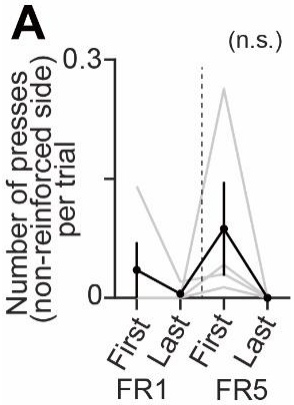

Figure 1—figure supplement 1. The number of non-reinforced lever presses showed no significant changes in a session-by-session comparison.

(A) Session-by-session comparison of the number of lever presses on the right 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 4 mice each, one-way repeated measures ANOVA, $F_{(3,9)}$ = 1.24, p = 0.35). Gray line, individual mouse. Black line, mean ± SEM.

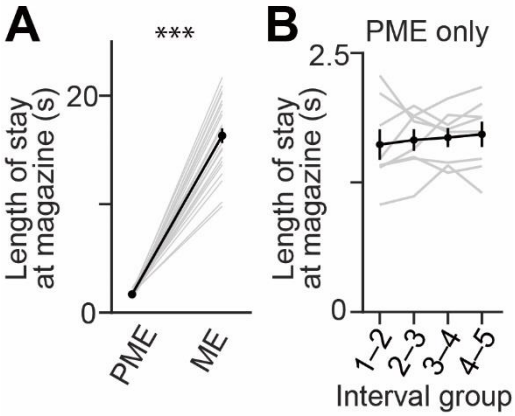

Figure 2—figure supplement 1. Mice stayed similar length of time at the food magazine following a PME regardless of the inter-press interval group.

(A) The mice stayed longer at the food magazine after an ME than after a PME (n = 24 sessions each from 4 mice, 6 sessions from each mouse; PME, 1.68 ± 0.05 s; ME, 16.31 ± 0.67 s; paired t-test, $t_{(23)} = -21.77$, $p = 7.5 \times 10^{-17}$). Gray line, individual session. Black line, mean ± SEM.

(B) No significant differences in the length of stay at the magazine following PMEs in the inter-press interval groups (n = 4 interval groups with 8 sessions each from 4 mice, FR5-First and FR5-2 sessions from each mouse, one-way repeated measures ANOVA, $F_{(3,21)} = 0.52$, $p = 6.8 \times 10^{-1}$). Gray line, individual session. Black line, mean ± SEM.

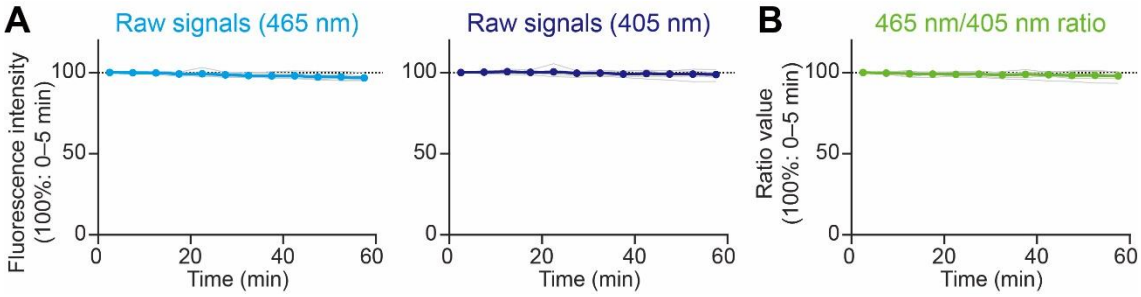

Figure 3—figure supplement 1. Stable fluorescence intensity for an hour enabled stable *in vivo* DA measurements from freely behaving mice.

(A) 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.

(B) Same as (A) 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 $p > 0.99$).

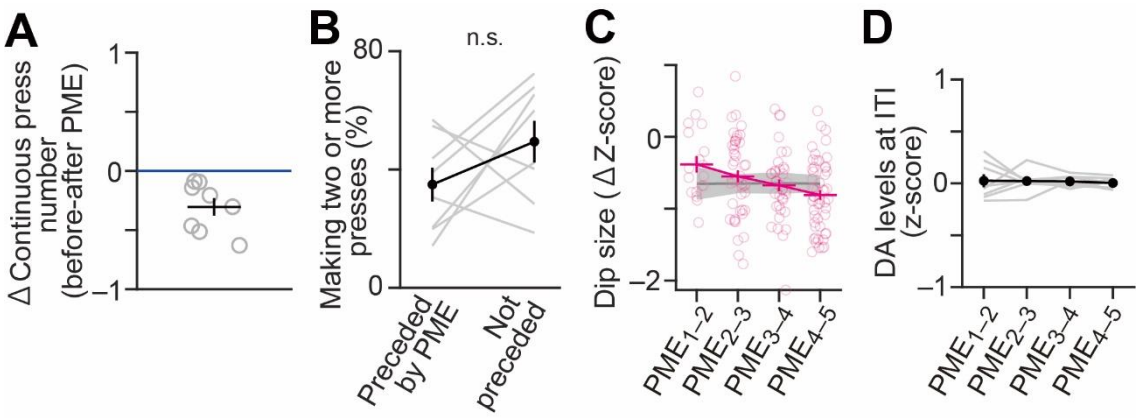

Figure 4—figure supplement 1. Analyses of behavior and DA dynamics upon PMEs.

(A) A PME induced general disappointment rather than a regret—comparison within a trial. Regret has been distinguished from general disappointment in previous studies (Steiner and Redish, 2014; Zeelenberg et al., 2010) even though both regret and disappointment are recognized when the reality is worse than expected. One feels regret if the reason for the present situation being unfavorable can be associated with his/her having made a bad choice—PME for mice in the current task—at the preceding decision, and if an alternative choice would make the present situation better. Our task did not provoke regrettable situations, because it did not deprive the mice of an expected reward in any trial even if the mice made one or multiple PMEs, corresponding to bad choices, up to the completion of the ratio requirement. Nevertheless, we did not exclude the possibility of the mice feeling regret following a PME. One feels more intense regret if less effort has been expended toward a goal (Dijk et al., 1999) and therefore wants a second chance to undo the regrettable event (Zeelenberg et al., 2010). Further, regret affects an animal's subsequent action in that it does what it did not do in the previous situation (Steiner and Redish, 2014). In our task, a PME might have made the mice regret not making a greater number of presses (efforts) prior to the PME. We focused on two adjacent press blocks within a single trial, separated by a PME, and compared the numbers of presses in these blocks. If this hypothesis was the case, the mice should have made more continuous presses (press block) after the PME than before the PME in a given trial. However, the latter press block consisted of a rather smaller number of presses (n = 8 sessions from 4 mice, FR5-First and FR2 sessions from each mouse, one-sample t-test, –0.31 ± 0.07 more presses after a PME compared to these before the PME, $t_{(7)} = 4.15$, $p = 4.3 \times 10^{-3}$), suggesting that the mice did not intend to undo the preceding potentially regretful event.

(B) A PME induced general disappointment rather than regret—comparison among trials. To clarify the effects of regret on behavior after excluding the effects of rising expectations with increasing proximity to task completion, we focused on lever presses with the same press number in different trials such as the second, third, or fourth presses. We hypothesized that the mice would make more efforts—make two or more continuous presses more often from a focused press—when the press was preceded by a potentially regret-inducing PME (Steiner and Redish, 2014). Contrary to this hypothesis, the presence of a preceding PME did not affect the percentage of the mice's subsequently making continuous presses (n = 8 sessions each from 4 mice, FR5-First and FR2 sessions from each mouse, preceded by PME, 34.9 ± 5.8% of the total case; not preceded by PME, 49.4 ± 7.0%; paired t-test, $t_{(7)} = 1.42$, p = 0.20), suggesting that

PMEs did not alter the mice's behavior in a way that regrettable events would do. These results confirmed that PMEs did not induce regret but induced general disappointment.

(C) Statistical analysis of DA dip amplitude in Fig. 4F plotted by PME group (one-way ANOVA, $F_{(3,150)} = 3.87$, $p = 1.1 \times 10^{-2}$). Each open circle represents an individual event. Magenta line, mean ± SEM. Gray line and shaded area, mean ± 95% confidence interval of shuffled data.

(D) DA levels during ITI periods (–1 to 0 s to sound cue onsets) did not show significant variances among PME group numbers (n = 8 sessions each from 4 mice, the first and second FR5 sessions from each mouse, one-way repeated-measures ANOVA, $F_{(3,21)} = 0.08$, $p = 0.97$).

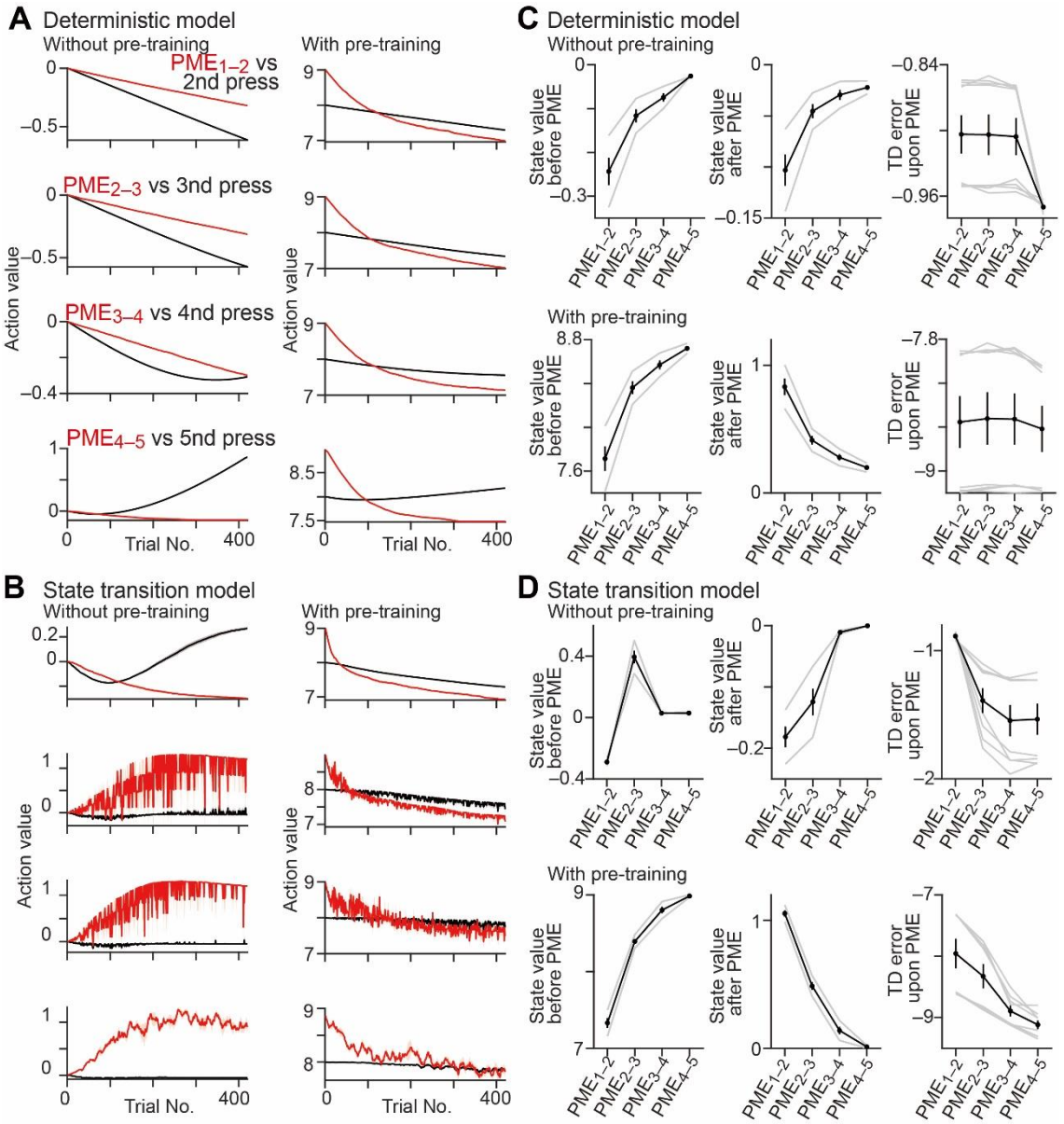

Figure 5—figure supplement 1. Trial-by-trial comparison of action value changes and comparison of average state or action values among PME groups.

(A) Comparison of trial-by-trial updates of action values plotted by the pair of actions that the agents choose from (e.g., second press versus $PME_{1–2}$) in the deterministic model without pre-training (left) and deterministic model with pre-training (right).

(B) Related to Fig. S5A but in the state transition model without pre-training (left) and the state transition model with pre-training (right).

(C) Plots of state values before PMEs (left), state values after PMEs (middle), and TD errors upon PMEs (right) in the first two sessions plotted according to PME group. Upper row: Results in the deterministic model without pre-training. Lower row: Results in the deterministic model with pre-training. Gray lines, individual data. Black lines, mean ± SEM. (n = 4 PME groups with 8 sessions each from 4 agents, first and second sessions from each agent, one-way repeated-measures ANOVA, deterministic model without pre-training, $F_{(3,27)} = 65.32$, $p = 1.7 \times 10^{-12}$ [state value before PME], $F_{(3,27)} = 56.56$, $p = 9.1 \times 10^{-12}$ [state value after PME], $F_{(3,27)} =$

15.6, p = 4.4 × 10^{-6} [TD error upon PME]; deterministic model with pre-training, $F_{(3,27)}$ = 154.41, p = 4.2 × 10^{-17} [state value before PME], $F_{(3,27)}$ = 205.99, p = 1.0 × 10^{-18} [state value after PME], $F_{(3,27)}$ = 8.17, p = 5.0 × 10^{-4} [TD error upon PME])

(D) Related to Fig. S5C but results in the state transition model without pre-training (upper row) or results in the state transition model with pre-training (lower row). Gray lines, individual data. Black lines, mean ± SEM. (n = 4 PME groups with 8 sessions each from 4 agents, first and second sessions from each agent, one-way repeated-measures ANOVA, state transition model without pre-training, $F_{(3,27)}$ = 203.64, p = 1.2 × 10^{-18} [state value before PME], $F_{(3,27)}$ = 82.47, p = 1.0 × 10^{-13} [state value after PME], $F_{(3,27)}$ = 29.08, p = 1.3 × 10^{-8} [TD error upon PME]; state transition model with pre-training, $F_{(3,27)}$ = 1178.79, p = 1.0 × 10^{-28} [state value before PME], $F_{(3,27)}$ = 1842.35, p = 2.5 × 10^{-31} [state value after PME], $F_{(3,27)}$ = 45.57, p = 1.1 × 10^{-10} [TD error upon PME]).

Supplemental References

Dijk, W.W.v., Pligt, J.v.d., and 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:

Steiner, A.P., and Redish, D.A. (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

Zeelenberg, M., Dijk, W.W.v., S.R.Manstead, A., and Pligt, J.d. (2010). The Experience of Regret and Disappointment. *Cognition and Emotion* **12**, 221-230. Digital Object Identifier: 10.1080/026999398379727, PMID: