Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013

Cell

CellPress

Article

# A Unified Framework for Dopamine Signals across Timescales

HyungGoo R. Kim,[1,9,10,*] Athar N. Malik,[1,2,9] John G. Mikhael,[3,4] Pol Bech,[1] Iku Tsutsui-Kimura,[1] Fangmiao Sun,[6,7,8] Yajun Zhang,[6,7,8] Yulong Li,[6,7,8] Mitsuko Watabe-Uchida,[1] Samuel J. Gershman,[5] and Naoshige Uchida[1,*]

[1]Center for Brain Science, Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA
[2]Department of Neurosurgery, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA
[3]Program in Neuroscience, Harvard Medical School, 220 Longwood Avenue, Boston, MA 02115, USA
[4]MD-PhD Program, Harvard Medical School, 260 Longwood Avenue, Boston, MA 02115, USA
[5]Department of Psychology, Center for Brain Science, Harvard University, 52 Oxford Street, Cambridge, MA 02138, USA
[6]State Key Laboratory of Membrane Biology, Peking University School of Life Sciences, Beijing 100871, China
[7]Peking-Tsinghua Center for Life Sciences, Beijing 100871, China
[8]PKU-IDG/McGovern Institute for Brain Research, Beijing 100871, China
[9]These authors contributed equally
[10]Lead Contact
*Correspondence: hyunggoo.r.kim@gmail.com (H.R.K.), uchida@mcb.harvard.edu (N.U.)
https://doi.org/10.1016/j.cell.2020.11.013

## SUMMARY

Rapid phasic activity of midbrain dopamine neurons is thought to signal reward prediction errors (RPEs), resembling temporal difference errors used in machine learning. However, recent studies describing slowly increasing dopamine signals have instead proposed that they represent state values and arise independent from somatic spiking activity. Here we developed experimental paradigms using virtual reality that disambiguate RPEs from values. We examined dopamine circuit activity at various stages, including somatic spiking, calcium signals at somata and axons, and striatal dopamine concentrations. Our results demonstrate that ramping dopamine signals are consistent with RPEs rather than value, and this ramping is observed at all stages examined. Ramping dopamine signals can be driven by a dynamic stimulus that indicates a gradual approach to a reward. We provide a unified computational understanding of rapid phasic and slowly ramping dopamine signals: dopamine neurons perform a derivative-like computation over values on a moment-by-moment basis.

## INTRODUCTION

Dopamine has important roles in controlling learning, motivation, and movement. Understanding what information dopamine conveys is critical for determining how dopamine regulates various functions. One influential idea is that a phasic activity of midbrain dopamine neurons represents temporal difference (TD) reward prediction errors (RPEs) used in reinforcement learning algorithms (Schultz et al., 1997; Niv, 2009; Eshel et al., 2015; Starkweather et al., 2017). Response patterns that conform TD RPEs have been observed in a number of animal species and under different task conditions (Bayer and Glimcher, 2005; Clark et al., 2012; Watabe-Uchida et al., 2017), and the RPE hypothesis has greatly affected our understanding of dopamine functions. However, many of these experiments have employed relatively simple behavioral paradigms using discrete stimuli and outcomes. Whether the same principle applies in more complex contexts remains to be examined.

Several studies using animals that can move within an environment have shown that dopamine concentrations in the striatum ramp up over a timescale of seconds (Phillips et al., 2003; Roitman et al., 2004; Howe et al., 2013; Hamid et al., 2016; Berke, 2018; Mohebi et al., 2019; Engelhard et al., 2019). Some authors have argued that these slow dopamine fluctuations cannot be readily explained by TD RPEs and have alternatively proposed that they represent the value of the state (state value or motivational value), which increases as the animal approaches a reward location (Berke, 2018; Hamid et al., 2016; Howe et al., 2013). Furthermore, a recent study (Mohebi et al., 2019) concluded that these ramping activities are absent in the spiking activity of dopamine neurons in the ventral tegmental area (VTA) and that ramping dopamine signals arise from local modulation of dopamine axons in the striatum. However, more work is needed to determine (1) what mechanisms underlie generation of ramping dopamine signals and (2) what behavioral conditions cause ramping dopamine signals.

Theoretically, value is separable from RPE. TD RPE $(\delta_t)$ is defined by

$$\delta_t = r_t + \gamma \widehat{V}(S_{t+1}) - \widehat{V}(S_t),$$

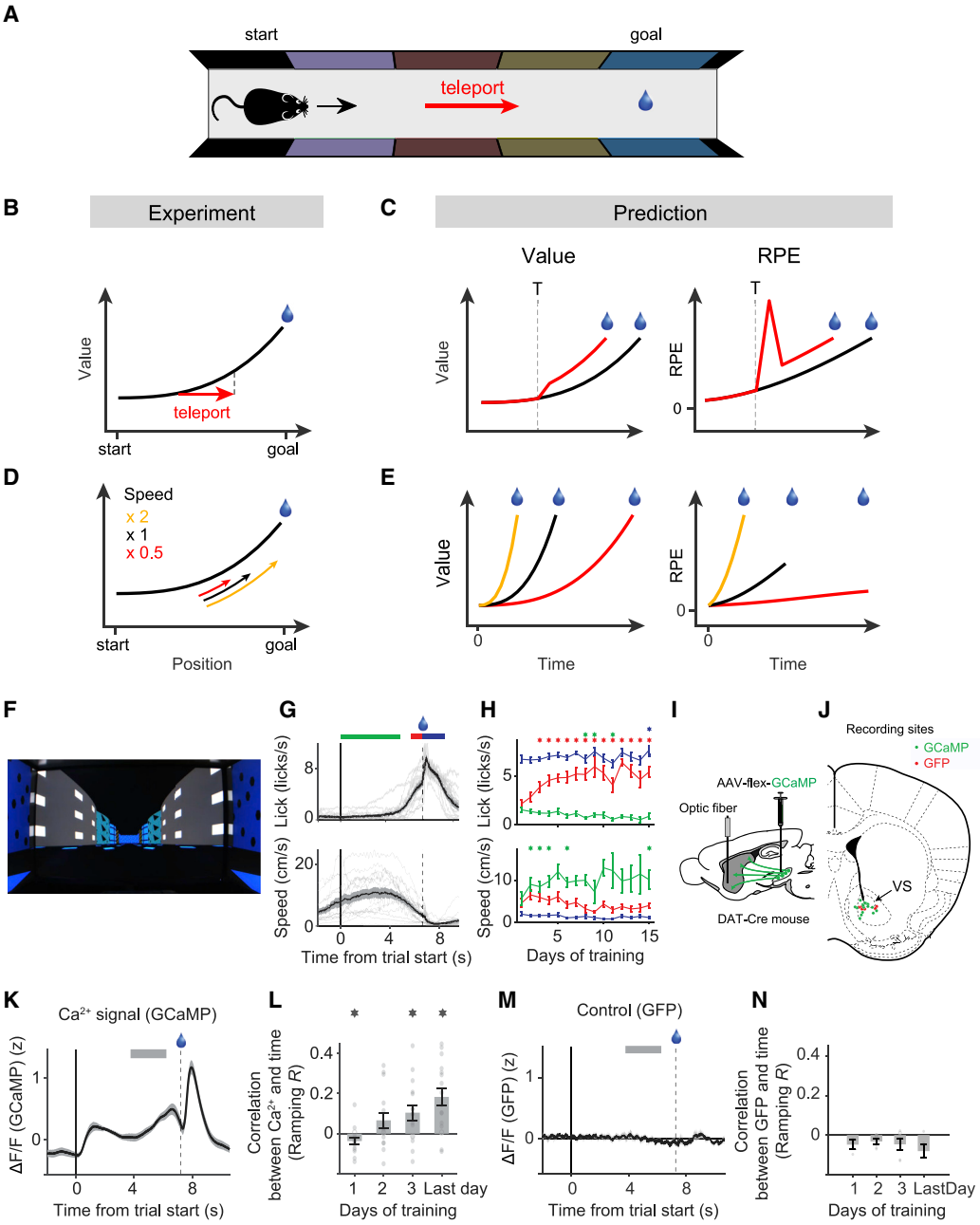



**Figure 1. Experiments to Dissociate Value and RPE Using Virtual Reality**

(A) The virtual linear track.

(B) State value as a function of position. Red arrow, teleportation.

(C) Predictions of how state value (left) and TD RPE (right) are modulated by teleportation (red curves).

(D) Speed manipulation.

(E) Predictions.

(F) An example scene at the starting position.

(G) Top: the time courses of lick rate (gray) and the average across animals (black) (n = 16 mice). Bottom: locomotor speed (gray) and the average (black). Green, red, and blue horizontal bars represent the time windows used for analysis in (H).

(H) Top: impulsive lick (green), anticipatory lick (red), and post-reward lick (blue) rates as a function of days of training. *$p < 0.05$ (n = 16 mice). Anticipatory lick increased, impulsive lick decreased, and post-reward lick did not change over days of training ($r = 0.39, -0.36, 0.04$; $p = 2.7 \times 10^{-7}$, $3.9 \times 10^{-6}$, 0.64, respectively; Spearman correlation). Bottom: locomotor speed.

(I) Fiber fluorometry (photometry) experiment.

(J) Recording locations in the experimental (green) and GFP control (red) animals (n = 16 and 6 mice, respectively).

*(legend continued on next page)*

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013



where $r_t$ is the reward the animal receives at time $t$, $S_t$ is the state the animal occupies at time $t$, $\gamma$ is the discounting factor $(0 < \gamma < 1)$, and $\widehat{V}(S_t)$ is the value of the state $S_t$ (i.e., state value), defined as the sum of all future rewards, where future rewards are discounted exponentially with factor $\gamma$ (STAR Methods). TD RPE contains terms that are approximately the difference between values at consecutive time points, $t$ and $t+1$ (i.e., $\gamma\widehat{V}(S_{t+1}) - \widehat{V}(S_t)$, where $\gamma$ is close to 1). Thus, in the absence of an immediate reward, TD RPEs are approximately the derivative of value. The idea that dopamine represents value is therefore incompatible with the view that dopamine represents TD RPEs.

Under many conditions, however, it is difficult to disambiguate RPE and value. A dopamine ramp can occur regardless of whether dopamine represents RPEs or value (Gershman, 2014; Lloyd and Dayan, 2015; Morita and Kato, 2014). A theoretical study showed that the shape of the value function matters; if the value function is a sufficiently convex function of proximity to reward, then a TD RPE can exhibit a positive ramp (Gershman, 2014) (STAR Methods; Figure S1). Therefore, the mere presence of a dopamine ramp does not distinguish the two possibilities.

Here we sought to develop experimental paradigms that empirically dissociate RPE from value. We focused on the core property of RPE, that RPE is approximately the derivative of value. Our experiments using visual virtual reality allowed us to tease apart these two possibilities. The results demonstrate that ramping dopamine signals are consistent with TD RPE but inconsistent with value.

## RESULTS

### Using Virtual Reality to Dissociate RPEs from Values

Imagine that a mouse moves along a linear track to obtain reward (Figure 1A). One can assume that the value of the animal's location increases monotonically as it approaches the reward. Now imagine that, while moving, the animal is suddenly teleported to a location closer to the goal (Figure 1B). If dopamine represents value, then it should exhibit a step-like increase at the time of teleportation and then continue increasing gradually, with the maximum level reached at the goal (Figure 1C, left). In contrast, if dopamine represents RPE, then it should exhibit a phasic excitation at the time of teleportation, reflecting an instantaneous increase in value (Figure 1C, right). Next, imagine that the speed of the mouse is manipulated (Figure 1D). If dopamine represents RPE, then the magnitude of the ramp will be modulated by the speed, with greater magnitudes for faster speeds (Figure 1E, right). In contrast, the value will reach the same level just prior to reward irrespective of the speed (Figure 1E, left). Importantly, this experiment directly tests the property of ramping itself, whether the ramp is consistent with RPE or value. The primary goal of these experiments is to distinguish whether dopamine signals are consistent with a monotonically increasing function that is dependent on the position or the derivative of that function. The former would support the value hypothesis, whereas the latter would support the RPE hypothesis.

We used virtual reality in head-fixed mice (Dombeck et al., 2007) to perform teleportation and speed manipulations. In the first set of experiments, the visual scene (Figure 1F) moved at a constant speed and the mice received a drop of water (5 μL) at the goal location (Video S1). Over several days, mice developed anticipatory licking near the goal (Figure 1G, top; n = 16 mice, p = 0.00043, Wilcoxon signed-rank test). Although the scene moved constantly irrespective of locomotor movement (i.e., "passive" condition), more than half of the animals developed running behavior (Figure 1G, bottom).

We first monitored calcium signals from dopaminergic axons projecting to the ventral striatum (VS, or *nucleus accumbens* core) (Babayan et al., 2018; Menegas et al., 2017, 2018) using fiber fluorometry (photometry) (Figures 1I and 1J). After training, axonal calcium signals ramped up progressively over a timescale of 3–4 s (Figures 1K and 1L; Figures S2E–S2H). We quantified the ramping based on correlation coefficients (r, "ramping *R*") between calcium signals and time (n = 16 mice, r = 0.18 ± 0.04; Pearson correlation using the average calcium trace was 0.45, 95% confidence interval [CI] = [0.43, 0.48], $p < 10^{-20}$). Across sessions, neither anticipatory licking nor running speed was correlated with the ramping signal (p = 0.37 and 0.13 for anticipatory licking and running speed, respectively; analysis of covariance [ANCOVA], n = 93 sessions from 16 mice). We did not observe a significant difference in ramping *R*s between slow- and fast-running animals (Figure S2J).

These ramping dopamine signals are unlikely to be a correlate of licking or motion artifacts. First, mice expressing a calcium-insensitive green fluorescent protein (GFP) did not exhibit ramping (Figures 1M and 1N). Second, we have not observed ramping signals using similar techniques in different tasks (Babayan et al., 2018; Menegas et al., 2017, 2018), although anticipatory licking was similar (also see delayed-reward task below). In addition to ramping, we observed a phasic response at the onset of scene movement and a slight decrease just before reward (STAR Methods).

### Ramping Dopamine Signals Are Consistent with RPEs

We then performed a set of 4 experiments to determine whether dopamine signals represent RPE or value. In experiment 1, in addition to the standard condition, we randomly interleaved three test conditions, which included a long teleportation, a short teleportation, or a 5-s pause (Figure 2A; Video S2). If dopamine represents value, then dopamine signals would show a step-like increase, arriving at the same level after a long and short teleportation, and should always reach the maximum level just before reward (Figure 2B, left). If dopamine represents RPE, then dopamine signals would show a phasic excitation whose magnitude scales with the length of teleportation (Figure 2B, right). Value depending on the distance to reward will stay constant when scene movement is paused, whereas RPE will

(K) Average axonal calcium signals (n = 16 mice). A gray horizontal bar depicts a temporal window used to compute Pearson correlations (ramping *R*).
(L) Ramping *R*s. *p < 0.05.
(M and N) Signals (M) and ramping *R*s (N) from GFP animals (p > 0.05, Wilcoxon signed-rank test for each day, n = 6 mice).
See also Figure S2.

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013



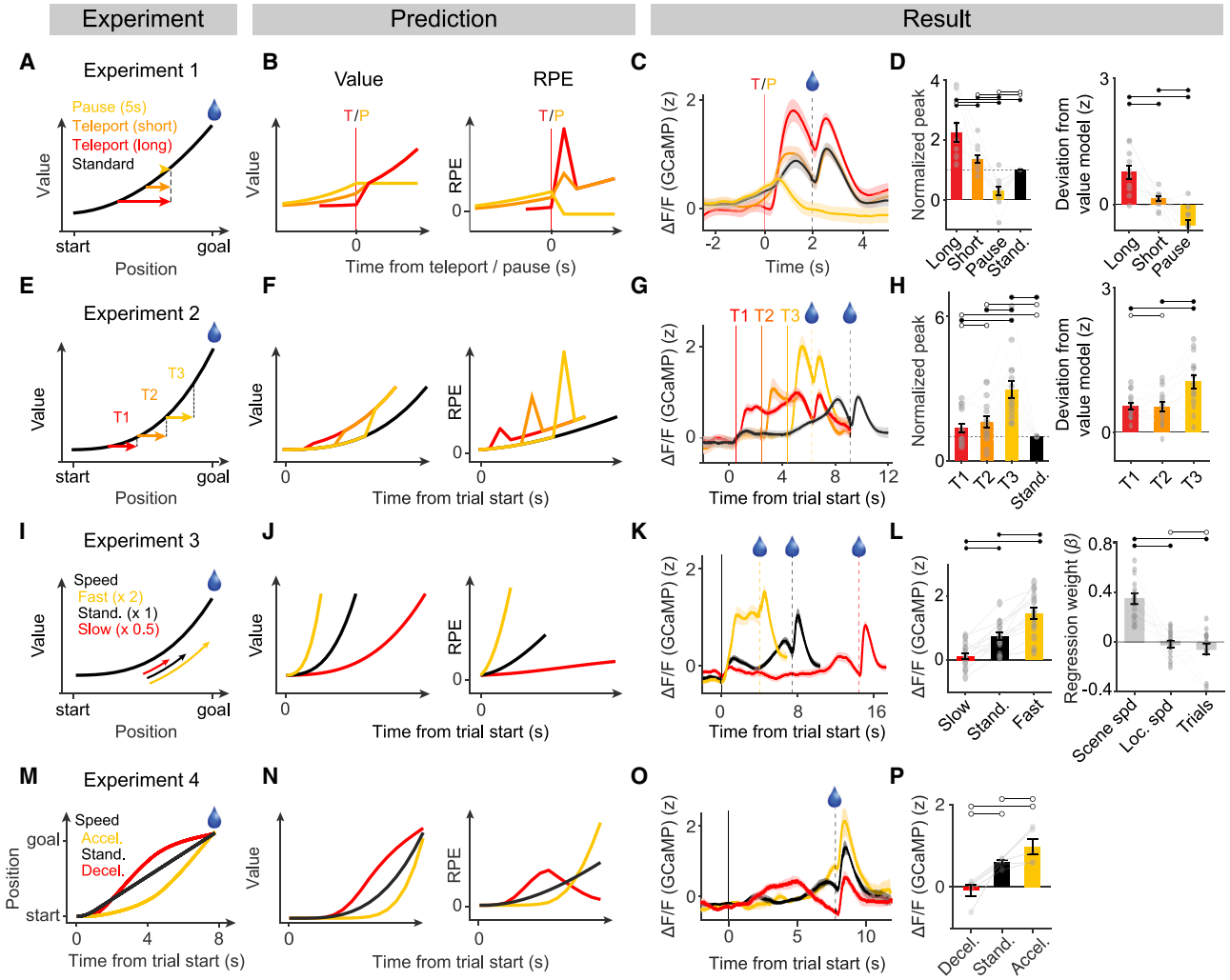



**Figure 2. Dopamine Axon Activities in the VS Are Consistent with RPE**

(A) Experiment 1. Long teleportation, short teleportation, and pause are depicted on the value function.

(B) Predictions. T, teleportation. P, pause.

(C) Average calcium signals aligned by teleportation or pause (n = 11 mice). Format as in (B). The trace of the standard condition (black) was aligned by reward onset.

(D) Comparisons of normalized peak responses (left) and residuals from the state value prediction (right) (n = 11 mice; Figures S4A–S4D). Horizontal bars with filled circles represent significant differences.

(E) Experiment 2. Teleportation at three positions (T1, T2, and T3).

(F) Predictions.

(G) Average calcium signals (n = 11 mice). Four mice whose scene speed was slightly faster than the rest of animals were excluded in the time course plots but included in other analyses (STAR Methods).

(H) Left: normalized peaks increase with proximity to the reward (median test $R = 0.45$, $p = 6.1 \times 10^{-5}$, n = 15 mice). Right: residuals from the state value prediction (median test $R = 0.20$; $p = 0.0031$, n = 15 mice).

(I) Experiment 3.

(J) Predictions.

(K) Average calcium signals (n = 15 mice).

(L) Left: Comparison of average pre-reward responses at [–1 s 0 s] relative to reward. Right: comparison of regression coefficients. The median of regression coefficients is positive only for the speed of scene movement ($p = 6.1 \times 10^{-5}$, 0.64, and 0.45, respectively; n = 15 mice).

(M) Experiment 4.

(N) Predictions.

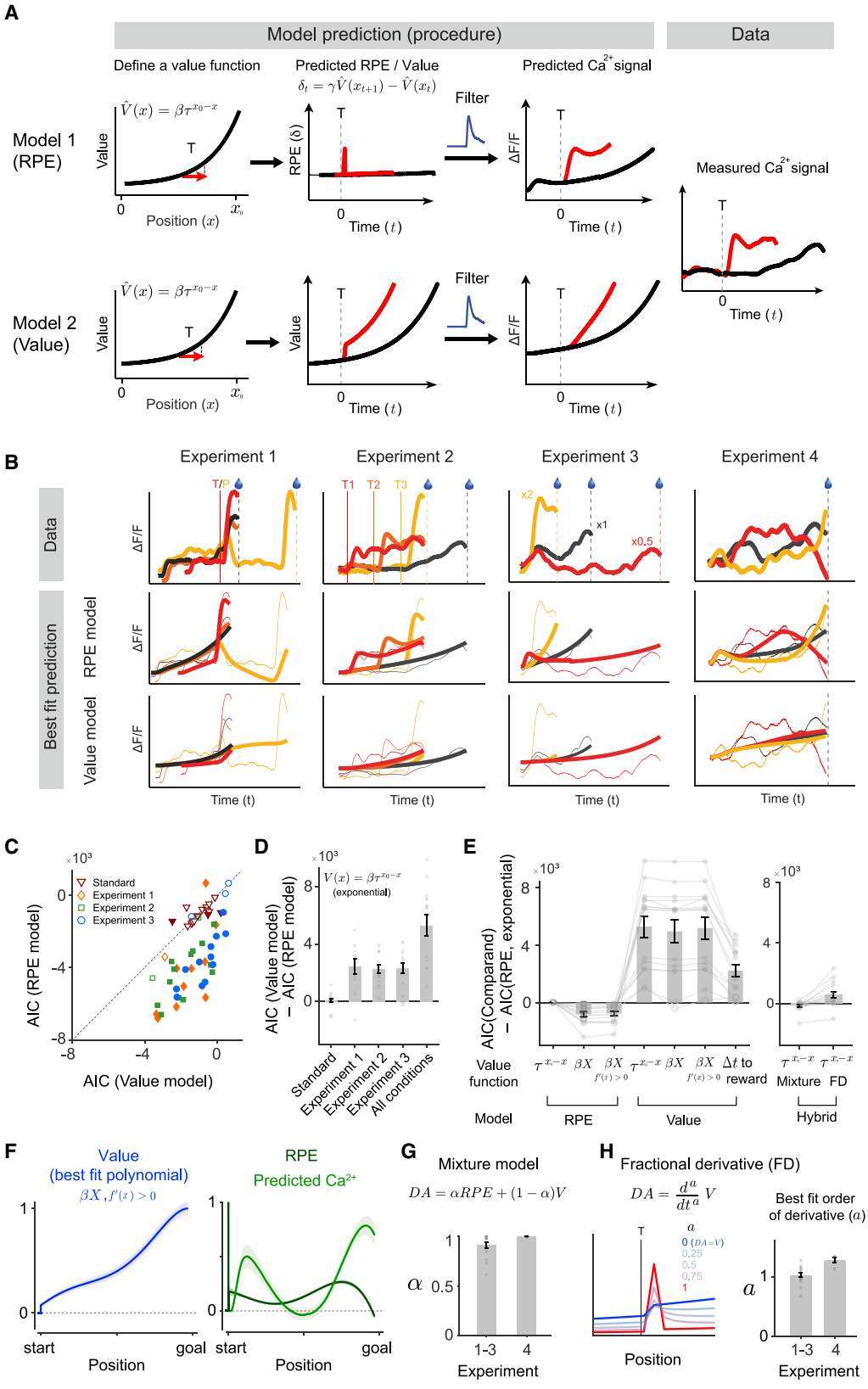
(O) Average calcium signals (n = 5 mice).

(P) Comparison of calcium signals before reward.

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013








*(legend on next page)*

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013



decrease to baseline because there is no change of value in time. There is some ambiguity regarding how "value" may behave under the pause condition; for example, if the animal judges that the task is aborted at the time of pause, then the value may also decrease to baseline. We used the results holistically to judge which hypothesis parsimoniously accounts for the entire data.

In teleportation trials, anticipatory licking and changes in locomotor speed reflected the destinations of teleportation (Figure S3B), confirming that mice used visual cues to predict reward rather than merely relying on elapsed time. A long teleportation evoked a large calcium transient whose peak was greater than the peak of the ramp under the standard condition (Figures 2C and 2D, left; ratio between the peaks, 2.25 ± 0.31; p = 0.0010, n = 11 mice; Figure S3A). The phasic excitation evoked by a short teleportation was smaller than that evoked by a long teleportation but was still greater than the peak of the ramp under the standard condition, violating the value hypothesis (Figures 2C and 2D, left; ratio, 1.35 ± 0.14; p = 0.024, n = 11 mice). In pause trials, the calcium signal decreased to the baseline level, followed by a phasic excitation when scene movement resumed (Figures 2C and 2D, left), consistent with RPEs.

To quantify these results, we generated predicted responses based on the value hypothesis (Figures S4A–S4D; STAR Methods). If dopamine represents value, then deviation from these predictions should be small and unsystematic. In most of the animals (9 of 11), the deviations of the observed signals followed a systematic pattern, supporting the RPE hypothesis (Figure 2D, right; median test *R*, r = – 0.64, n = 11 mice, p = 0.002; see STAR Methods for the definition of test *R*).

Because value is unobservable, it is generally difficult to assess the shape of value function. In experiment 2, we sought to infer the shape of value function (Figures 2E–2H; Figures S3C and S3D). In test trials, mice were teleported forward from one of three locations by the same distance (Figure 2E). If the underlying value function has a convex shape, then the magnitude of response should be larger, with teleportation occurring at locations closer to the goal. Indeed, the phasic calcium signals followed this pattern (Figure 2H), consistent with a convex value function.

To test whether the ramping itself represents RPEs, we moved the scene either fast (×2 speed) or slow (×0.5 speed) in test trials (experiment 3; Figures 2I–2L; Figures S3E and S3F; Video S3). Observed calcium signals were consistent with the RPE predictions (Figures 2K and 2L, left; p = 6.1 × $10^{-5}$ and 6.1 × $10^{-4}$, n = 15 mice). A regression analysis indicated that the magnitude of ramping can be predicted by the speed of the scene but not by locomotion speed (Figure 2L, right). GFP control mice did not show systematic modulation (Figure S4F).

We note, however, that there was a sudden increase in the signal soon after onset of fast scene movement. This may be because the speed of scene movement became a cue predictive of an early reward. Although this is still consistent with RPE, we designed an additional experiment that minimized this potential confound (experiment 4; Figures 2M–2P; Figures S3G–S3I). The speed of scene movement was modulated dynamically over time (Figure S3I). This allowed us to change the speed without altering the time to reward between conditions. We found that dopamine responses before reward changed according to the instantaneous speed (Figure 2O). The calcium signals immediately preceding the goal diverged greatly (Figure 2P; p = 0.002, df = 2, n = 5 mice, F = 14.3, one-way repeated-measures ANOVA), violating the value account, which predicts that dopamine signals should reach the same level at the goal regardless of the speed.

We next used a model fit analysis to test whether the data can be explained better by RPE or value. The state value was first defined as a function of space (Figure 3A). Based on this value function, we predicted calcium signals under each condition. We then obtained a set of parameters that minimized the residual sum of squares (Figures 3A and 3B). The goodness of fit was quantified using the Akaike information criterion (AIC) to penalize the number of parameters in a model. We first used a value function whose value is discounted by a fixed rate ($\tau$) as a function of the distance to the target (exponential value function; the requirement of a particular shape for the value function will be relaxed later). The RPE model explained the data far better than the value model under all experimental conditions (Figure 3D; p < 0.004 for all four fits with manipulated experiments; $H_0$, individual median ΔAIC is zero; n = 11, 15, 15, 15 for experiments 1, 2, 3, and all). In contrast, the difference was not significant under the standard condition (Figure 3D, standard; p = 0.07, n = 16 mice), indicating that our analysis is unbiased.

We further fitted the data with value functions of more arbitrary shapes (e.g., fifth-order polynomial) (Figure 3E, $\beta X$; Figure S4H), allowing us to derive a value function in a more data-driven manner. We also included a value model in which the state value was computed based on time to reward given the current speed (Δt to reward). The RPE model with a polynomial value function best explained the data. However, even the simple exponential

**Figure 3. RPE Models Explain the Data Better Than Value Models**

(A) Model fitting procedure. Blue curves, GCaMP filters.

(B) Fit examples. Top: the data. Center: best fit with the RPE model. Thick lines, model prediction. Thin lines, data. Bottom: best fit with the value model.

(C) Comparisons of AICs based on the exponential value function. Filled symbols, p < 0.05 (permutation test). A smaller AIC value indicates a better fit.

(D) Difference between the two models in (C).

(E) Left: AIC relative to the exponential RPE model. The combined dataset for experiments 1–3 was used. $\tau^{(x_0-x)}$, exponential discounting; $\beta X = \sum(\beta_k x^k)$, fifth-order polynomial ; $\beta X$, $f'(x) > 0$, fifth-order polynomial with the constraint of monotonical increase; Δt to reward, value based on time to reward given the current speed. Filled dots indicate significance. Right: hybrid models. Mixture, $(1-\alpha)V(x)+\alpha\delta(x)$; FD, fractional derivative model. Significance is not shown for the FD.

(F) The shape of value function (left), RPE (right, dark green), and the predicted calcium signal (right, green) obtained by the RPE model using $\beta X, f'(x) > 0$. The peak of the transient RPE at trial start is not shown.

(G) The optimal $\alpha$ in the mixture model.

(H) The best-fit order of derivative ($a$) in the FD model.

See also Figures S3 and S4.

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013





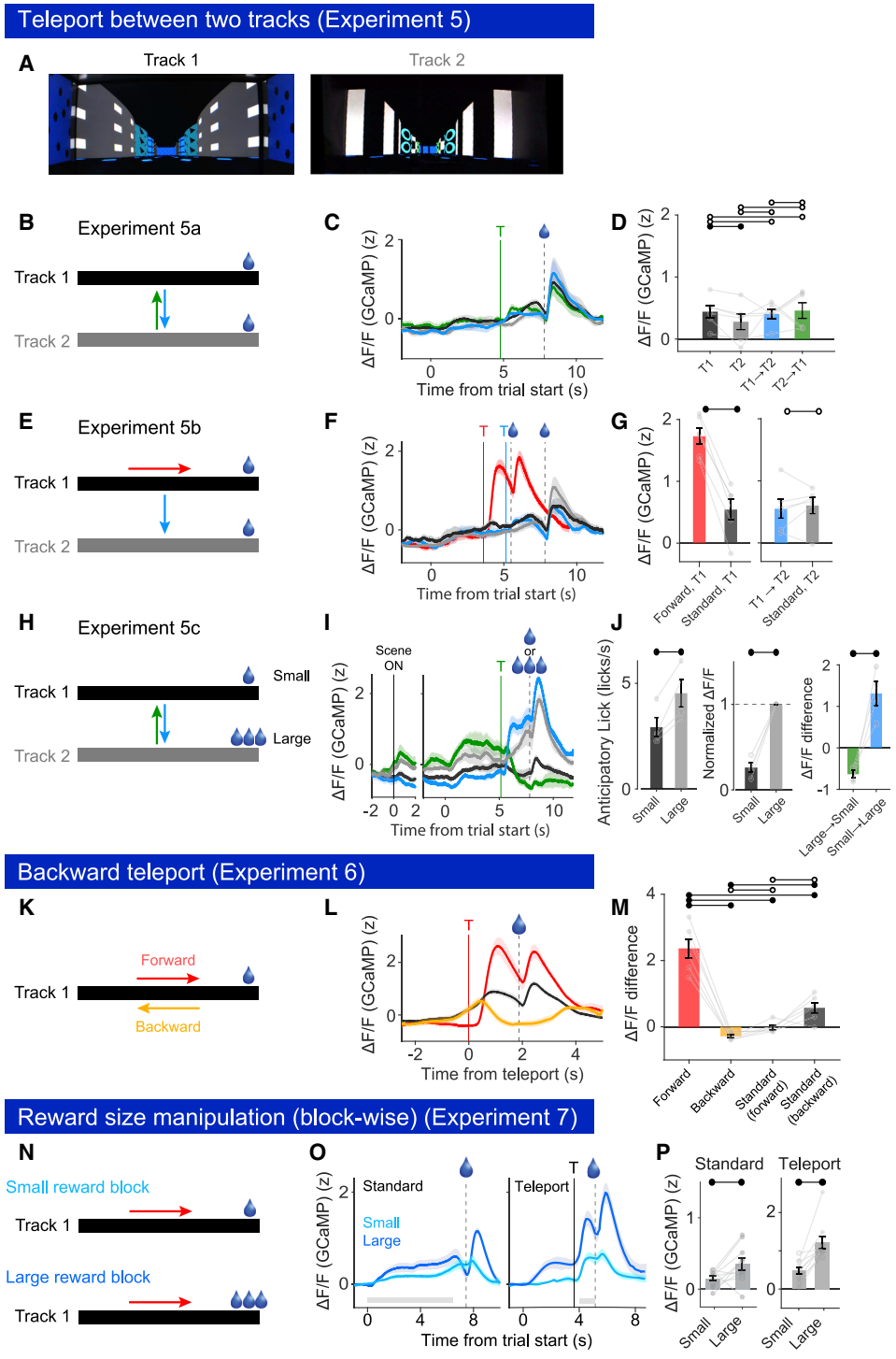



**Figure 4. Ramping and Teleportation Responses Cannot Be Explained by a Sensory Surprise**

(A) The scenes on tracks 1 and 2.

(B) Experiment 5a. Arrows, teleportation between tracks.

(C) Average calcium signals (n = 6 mice).

(D) Baseline-subtracted calcium responses.

(E) Experiment 5b. Red, forward teleportation; cyan, between-track teleportation.

(F) Average calcium signals (n = 6 mice).

(G) Baseline-subtracted calcium responses.

*(legend continued on next page)*

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013



RPE model outperformed all of the value models (Figure 3E; $p < 0.0003$; $H_0$, individual median ΔAIC is zero; n = 15 mice). These results indicate that the calcium signals in experiments 1–3 are better explained by RPE than by state value. Note that the fitted models also captured the initial transient response and the dip right before reward (see STAR Methods for a note regarding the shape of the fitted value function).

These analyses demonstrate that RPE models are better than value models when tested one against another. However, it is possible that the responses lie somewhere between these two possibilities. To address this, we first considered a linear combination of RPE and value, with the weight $\alpha$ $(0 \leq \alpha \leq 1)$ representing the fraction of RPE signals (a mixture model). The fit using this mixture model only barely improved compared with the RPE model (Figure 3E, right; the mean $R^2$ increased by 2%). The weight for the RPE term $(\alpha)$ was close to 1 (Figure 3G; experiments 1–3, 0.92 ± 0.12; experiment 4, $0.99 \pm 1.2 \times 10^{-4}$, mean ± SD). Second, we considered the possibility that the responses are between value and RPE in terms of the order of the derivative. Specifically, the RPE approximates the first-order derivative of the value function $(dV/dt)$, whereas the value function itself is its own zeroth-order derivative. The method of "fractional derivatives" allows one to define a non-integer order of derivative $(d^a/dt^aV)$ (Podlubny, 1998) by which one can obtain a gradual transition between value and RPE by varying *a* from 0 to 1 (Figure 3H, left). We found that the best-fit order of derivative obtained from the data using an exponential value function was close to 1 (Figure 3H, right; experiments 1–3, 1.1 ± 0.12; experiment 4, 1.28 ± 0.08, mean ± SD).

These results demonstrate that the RPE model, which computes the first-order derivative of the value function, is a superior model to explain the dopaminergic axonal activity in the VS, with little contribution of value.

## Dopamine Axons in the VS Do Not Respond to Sensory Surprise

Some recent studies have suggested that dopamine neurons are activated by sensory surprise, sensory (identity) prediction error, or arousal (Schultz, 2019; Stalnaker et al., 2019; Takahashi et al., 2017). We next tested whether the above responses were due to sensory surprise using teleportation between two tracks (Figure 4A). In test trials, mice were teleported between the tracks without changing the distance to the goal so that a teleportation event caused a sensory prediction error without causing a change in value (Figures 4B–4D; Video S4). We did not observe a transient excitation at the time of between-track teleportation (Figure 4D; Figure S4I; p = 0.31 and 0.84, n = 6 mice), although forward teleportation caused a large transient activation (Figures 4E–4G; Figure S4J). The lack of response during between-track teleportation is not due to failure to distinguish the two tracks or failure to recognize teleportation. When different amounts of reward were assigned to the two tracks (Figures 4H–4J; Figure S4K), we observed different levels of anticipatory licking and the calcium signal between the two tracks (Figure 4J, left and center; p = 0.019 and 0.001 for licking and calcium signal, respectively; n = 4 mice, paired t test). Furthermore, between-track teleportation caused a transient change in the calcium signal consistent with the change in the state values (Figure 4J, p = 0.012, n = 4 mice, paired t test). Finally, we also performed backward teleportation with the same magnitude as forward teleportation. Although the amount of sensory surprise was plausibly similar between these conditions, backward teleportation caused a decrease rather than an increase in the calcium signal (Figures 4K–4M; Figure S4L). These results indicate that a pure sensory surprise does not excite dopamine neurons but that a change in value is important.

We next examined whether the magnitudes of ramping and teleportation responses are sensitive to reward magnitudes. The amount of reward in track 1 was altered across blocks of trials (Figures 4N–4P). In large-reward blocks, mice showed greater anticipatory licking compared with small-reward blocks (p = 0.008, n = 10 mice). The magnitudes of ramping as well as phasic response were greater in large-reward blocks (Figures 4O, left, and 4P, left; p = 0.049, n = 10 mice; Figures 4O, right, and 4P, right; p = 0.0020, n = 10 mice, respectively). Thus, ramping and transient responses to teleportation are sensitive to outcome values.

The responses observed in our experiments cannot be explained by sensory surprise but can be explained parsimoniously by TD RPE—tracking changes of value.

## The Spiking Activity of Dopamine Neurons Exhibits Ramping Consistent with RPE

The above results indicate that the activity of dopamine axons in the VS is consistent with TD RPEs. However, it remained unclear whether these results held at the single-neuron level. For instance, different populations of dopamine neurons may separately underlie ramping, transient responses, and speed-dependent modulations. Furthermore, a recent study concluded that the spiking activity of VTA dopamine neurons does not ramp

(H) Experiment 5c. Arrows, between-track teleportation. A large reward was given in track 2.
(I) Average calcium signals (n = 4 mice).
(J) Left: comparison of anticipatory licking (3 of 4 mice showed a significant difference; Wilcoxon rank-sum test using trial data). Center: comparison of calcium responses (3 of 4 mice showed a significant difference; unpaired t test using trial data). Right: transient changes of calcium responses at teleportation (p = 0.006 and 0.021, large to small and small to large, respectively; n = 4 mice, paired t test).
(K) Experiment 6. Arrows, forward (red) and backward (orange) teleportation.
(L) Average calcium signals (n = 6 mice).
(M) Comparisons of calcium responses. Responses to the forward teleportation were significantly larger than responses to backward teleportation (p = 0.03, n = 6, Wilcoxon signed-rank test).
(N) Experiment 7. The reward size was altered across blocks of trials.
(O) Average calcium signals (n = 10 mice).
(P) Comparison of calcium responses, quantified using the time windows depicted in (O) (gray bars). Left: ramp magnitudes. Right: teleportation responses.
See also Figure S4.

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013



CellPress

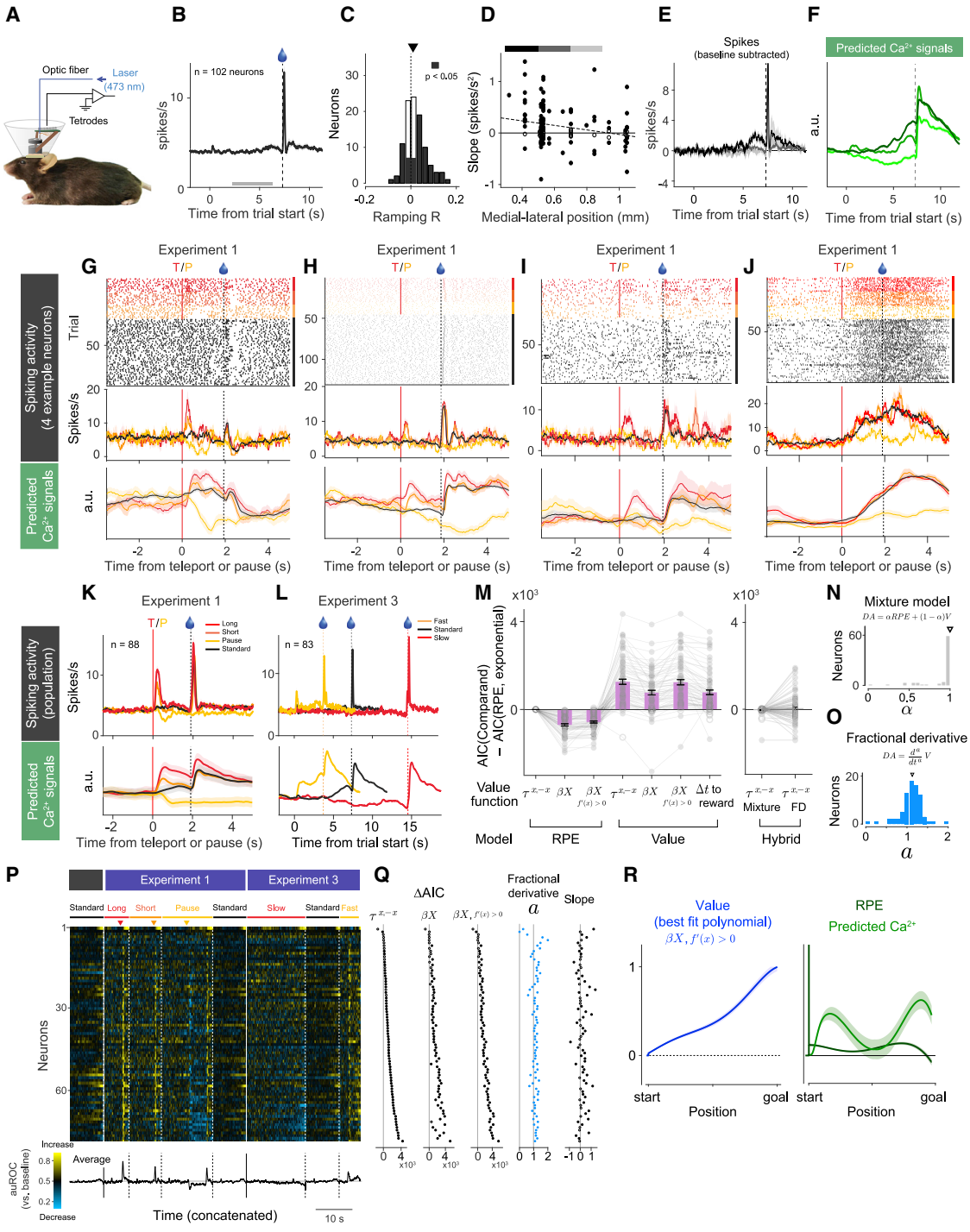



**Figure 5. Spiking Activity of VTA Dopamine Neurons Accounts for the Ramping Calcium Signals**

(A) Experiments.

(B) Average firing rates of VTA dopamine neurons (n = 102) under the standard condition. Gray bar, a time window used to quantify ramping in (C).

(C) Distribution of ramping *R*s. The median (triangle) is positive (p = 0.0001, n = 102 neurons).

(D) Ramping slope as a function of ML locations (n = 122). Gray bars, subgroups of neurons used in (E) and (F) (black, n = 16 neurons from 3 mice; dark gray, n = 66 neurons from 4 mice; gray, n = 20 neurons from 3 mice). The median slope was greater than zero in the two medial groups (p = 0.004, 0.009, and 0.39, respectively). Dashed line, type 2 regression fit.

(E) Average firing of groups of neurons indicated in (D).

(F) Calcium signals predicted from spikes. Darkness indicates the groups in (D).

*(legend continued on next page)*

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013



up, suggesting local modulation of dopamine axons independent of somatic spiking activity (Mohebi et al., 2019). To address these issues, we next characterized the spiking activity of VTA dopamine neurons (we focused on experiments 1 and 3, which contain the main experiments shown in Figure 1) (Figure 5A). The light-gated cation channel (channelrhodopsin-2) was expressed in dopamine neurons, and the recorded neurons were classified as dopaminergic based on the short-latency responses to laser pulses (Figures S5A–S5D; Cohen et al., 2012; Lima et al., 2009). Because VS-projecting dopamine neurons spread along the mediolateral (ML) axis in the VTA (Farassat et al., 2019), we recorded from a wide range along the ML axis (in total, 122 neurons from 13 mice; 102 neurons in the VTA [ML < 900 $\mu$m]).

Most of the optogenetically identified dopamine neurons showed a strong phasic response to reward (Figure 5B; 92 of 102 neurons show significance). On average, spiking activity exhibited a positive ramp (slope, 0.16 ± 0.04 spikes/s$^2$). Across neurons, we observed a significant diversity in ramping, with some neurons gradually ramping up and some others ramping down (Figure 5C; Figure S5E). In more medial recordings, more neurons showed a positive ramp, and the magnitude of ramps was greater (Figures 5D and 5E, black; slope, 0.41 ± 0.11 spikes/s$^2$). In more lateral recordings, fewer neurons showed a positive ramp, and the magnitude of ramps was smaller (Figures 5D and 5E, dark gray; slope, 0.12 ± 0.04 spikes/s$^2$) and diminished, on average, at the most lateral location (Figures 5D and 5E, gray; slope, 0.09 ± 0.09 spike/s$^2$). There was a significant correlation between the ramping slope and the ML position (Figure 5D; r = –0.21, p = 0.019, n = 122 neurons, Spearman correlation), consistent with single-neuron imaging results (Engelhard et al., 2019).

We next asked whether the small increase in spiking can account for the ramping in calcium signals (Figures 1 and2). Predicted calcium signals were computed based on the spiking activity and the impulse response of GCaMP6m (Figure S5P; STAR Methods). We found that ramping signals became more prominent and the phasic responses less prominent compared with the raw firing rates, resulting in a prominent positive ramp in the predicted calcium response from medial neurons and a slightly negative ramp from lateral neurons (Figure 5F).

It remains unclear whether the recorded neurons projected to the VS, where we monitored axonal calcium signals (Figure 2). To remedy this issue, we next performed fluorometry from the somata of VTA dopamine neurons that were labeled retrogradely from our recording site in the VS using a modified rabies virus. Inspection of the injection site and the location of somata of labeled neurons indicated that dopamine neurons that projected to our fluorometry recording sites were widely distributed in the VTA but enriched in the medial VTA (Figures S5H and S5I; Lammel et al., 2008; but see Ikemoto, 2007). Calcium signals recorded in these retrogradely labeled neurons exhibited a positive ramp (Figure S5J).

We next examined whether single-neuron activities were consistent with RPE or value using experiments 1 and 3. As discussed above, the RPE hypothesis would be supported by a phasic response at teleportation and changes in the magnitude of ramping in speed manipulations (Figures 1A–1E). In contrast, the presence of positive and negative ramps does not distinguish the two hypotheses; a sufficiently convex value function will cause a positive ramp, as discussed above, but a less convex value function can result in a negative ramp, based on the RPE hypothesis (Figure S1). We found that a large fraction of neurons, including those with a positive, negative, or no ramp (Figures 5G–5I), exhibited a phasic response at the time of teleportation (68 of 88 neurons). A phasic response was observed irrespective of whether they had a positive or negative ramp (r = 0.13, p = 0.24; n = 88 neurons, Spearman correlation between response to the long teleportation and ramping *R*). The magnitude of the ramp was modulated by the speed of scene movement (median test *R,* r = 0.18, p = 5.2 × 10$^{-12}$, n = 88 neurons; 27 of 88 neurons significantly greater than zero, Spearman correlation using trial data). The average spiking activity also showed features that are consistent with RPEs, and the predicted calcium responses resembled the axonal calcium signals in the VS (Figures 5K and 5L). We also observed a very small number of neurons that did not conform to RPEs (Figure 5J).

We next performed a model fit analysis on individual neurons. We applied the same fitting procedure (Figure 3A) after smoothing the spiking activity with a relatively narrow filter (Figures S5P, S5Q, and S5R). On average, an RPE model with a simple exponential value function outperformed all of the value models we

(G–I) Example neurons. An example neuron that showed positive ramp (G; ramping *R*, r = 0.018, p = 0.009), a negative ramp (H; ramping *R*, r = –0.0214, p = 0.0001), and no ramp (I; ramping *R*, r = 0.005, p = 0.49).
(J) A neuron that showed a value-like responses (the top row in P).
(K) Top: average spiking activity (n = 88). Bottom: predicted calcium signals.
(L) Top: average spiking activity (n = 83). Bottom: predicted calcium signals.
(M) Comparisons of goodness of fits (AIC). Left: ΔAIC relative to the RPE model based on the exponential value function. Format as in Figure 3E. Right: ΔAIC in hybrid models.
(N) The best-fit $\alpha$ for the mixture model.
(O) The best-fit order of derivative ($a$).
(P) Single-neuron activities. Neurons are sorted by ΔAIC between the value and RPE models (exponential value function) (n = 78 neurons). The area under the receiver operating characteristic curve (auROC) at each time bin was used to quantify firing rate changes from baseline. Arrowheads indicate the time of teleportation or pause onset.
(Q) Single-neuron activities. ΔAIC between the value and RPE models using three different value functions as in Figure 5M, the best-fit order of derivative in the FD, and the slope of ramping are shown (n = 78 neurons).
(R) Left: average of the normalized best-fit value functions (n = 78 neurons). Right: the average RPE predicted from the value functions of individual neurons (dark green) and predicted calcium signal (green).
See also Figure S5.

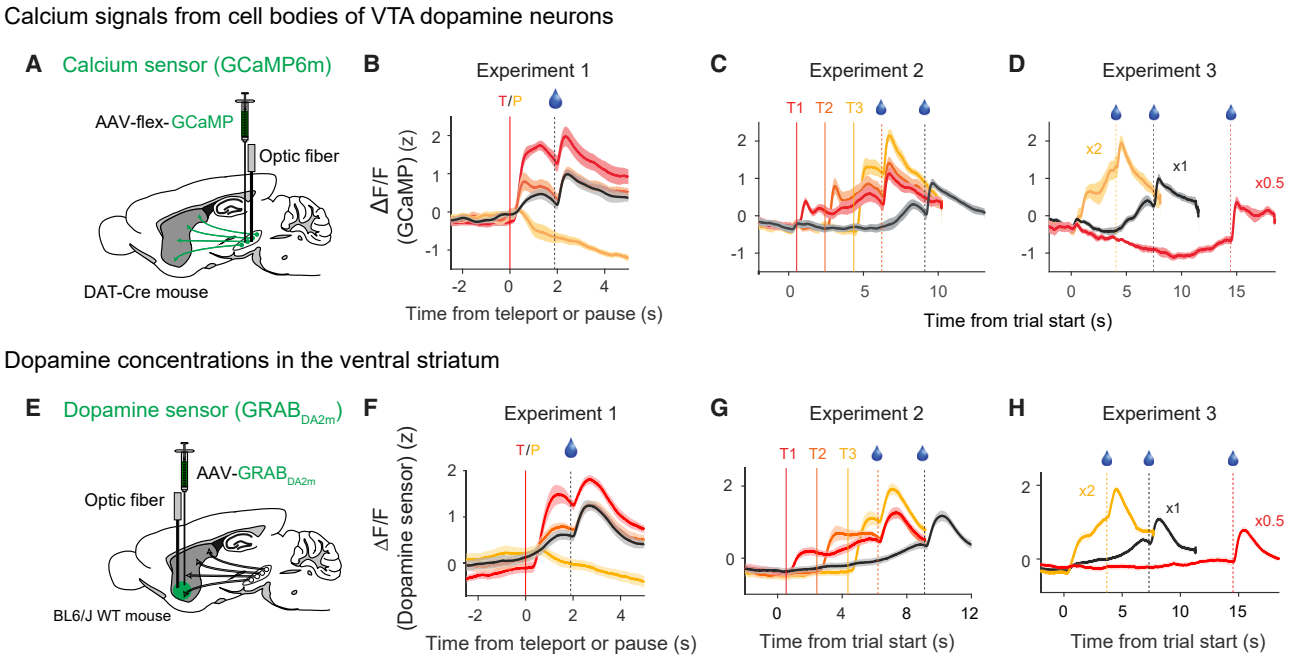



**Figure 6. Somatic Calcium of VTA Dopamine Neurons and Dopamine in VS Signal RPEs**

(A) Experiment.

(B–D) Average calcium signals in experiment 1 (B, n = 6 mice), experiment 2 (n = 6 mice), and experiment 3 (n = 5 mice).

(E) Experiment.

(F–H) Average dopamine signals in experiment 1 (F, n = 9 mice), experiment 2 (G, n = 10 mice), and experiment 3 (H, n = 10 mice).

See also Figure S6.

tested (Figure 5M, left; $p < 2 \times 10^{-10}$ for all value models; $H_0$, individual median ΔAIC is zero; n = 78). A mixture model that combined RPE and value barely improved the fit compared with the RPE model (Figure 5M, right; mean $R^2$ increase by 1.8%; AIC difference is significant in 1 of 78 neurons), and the weight for RPE was close to 1 (Figure 5N; $\alpha$ = 0.91 ± 0.20, mean ± SD). The best-fit orders of fractional derivative for individual neurons were close to 1 (Figure 5O; a = 1.11 ± 0.27, n = 78 neurons, mean ± SD). Thus, just as with axonal calcium signals, spiking activity encoded RPE nearly exclusively, with little contribution of value.

For a closer inspection, we sorted neurons based on how well the activity was fit by an RPE model over a value model (ΔAIC) (Figures 5P and 5Q). The phasic responses to teleportation, the dip at a pause, as well as the responses to different speeds were widespread across neurons (Figure 5P). A model fit analysis showed that most neurons preferred RPE over value models (Figure 5Q; significant for at least 62 of 78 neurons, three columns of ΔAICs, permutation test). This was true regardless of whether individual neurons ramped up or down (Figure 5Q; slope column, r = 0.09, p = 0.42, n = 78 neurons, Spearman correlation). We next inferred the shape of value functions through a model fit analysis using a flexible value function. The average value function derived from the spiking data, although performed completely independent of the measurement of axonal calcium signals, exhibited a remarkable resemblance to the value function derived from the axonal calcium signals (Figure 5R; compare with Figure 3F).

These results indicate that most of single neurons were consistent with RPE and relatively spread across the VTA. Among these neurons, those located in the medial region exhibited more positive ramps.

## Somatic Calcium in the VTA and Dopamine Concentration in the VS Are Consistent with RPEs

One potential caveat with the rabies virus (Figure S5J) is its toxicity. We therefore used an adeno-associated virus to express GCaMP in VTA dopamine neurons and recorded somatic calcium signals. We found that the calcium signals measured in the VTA showed a similar level of ramping. Further testing in experiments 1–3 (Figures 6A–6D; Figure S6) showed a striking similarity to the predicted calcium dynamics from the population of VTA dopamine neurons (Figures 5K and 5L bottom).

The dynamics of dopamine concentration can be affected by additional factors after release, such as diffusion, receptor binding, and reuptake. If the RPE signal is temporally integrated, then the dopamine concentration could be converted to a value-like quantity. Therefore, we measured dopamine concentration in the VS using a genetically encoded dopamine sensor ($GRAB_{DA2m}$) (Figure 6E). We found that the dopamine signals were similar to the calcium signals in other experiments (Figures 6F–6H). Transient excitation to long teleportation exceeded the peak of ramping under the standard condition (Figure 6F; p = 0.004, n = 9 mice), and the magnitude of ramping signals was modulated by the speed of scene movement (Figure 6H; median test *R*, r = 0.43, p = 0.002, n = 10 mice; 9 of 10 mice showed

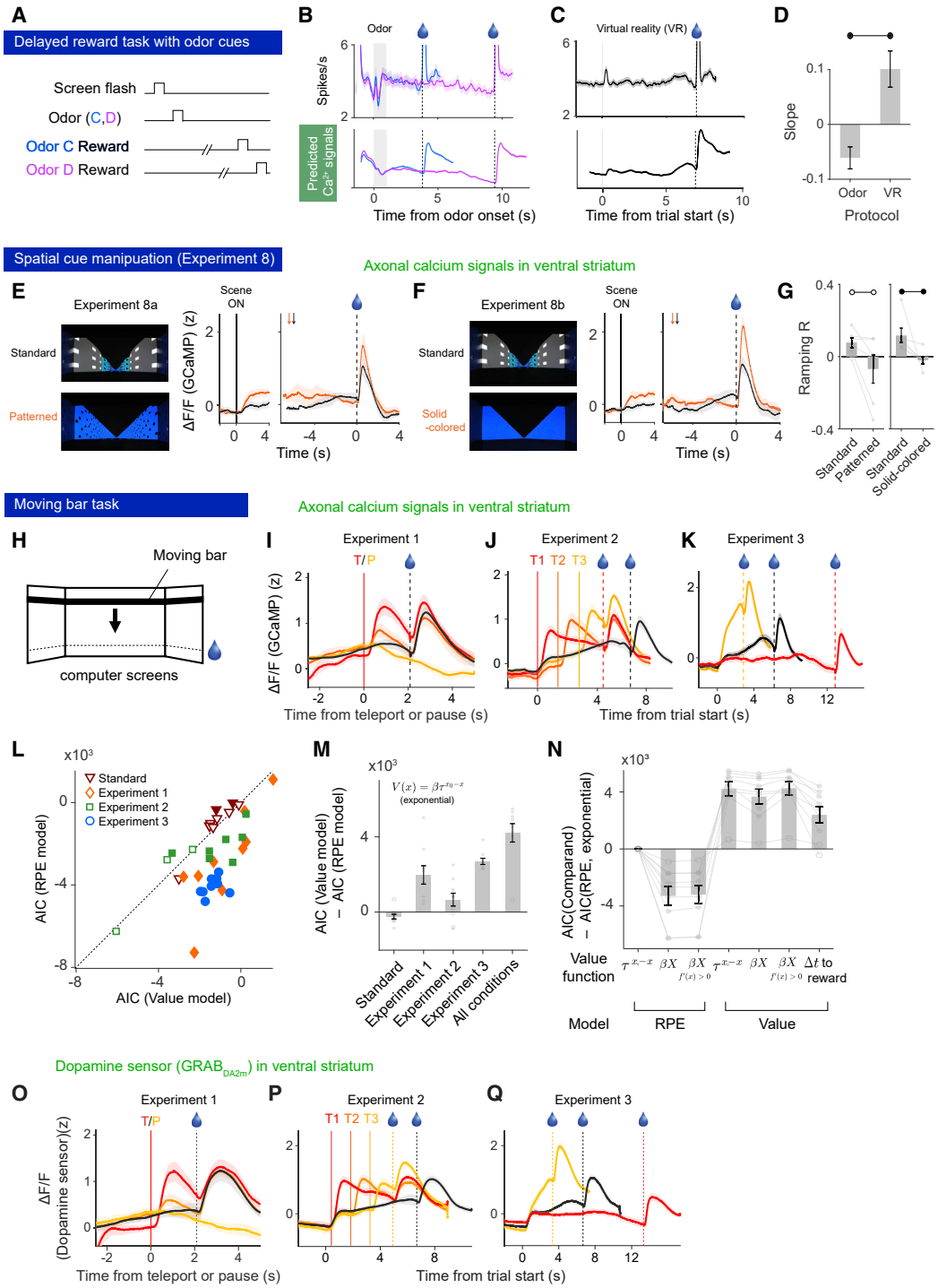



**Figure 7. Dynamic Sensory Stimulus Indicating Reward Proximity Can Cause a Dopamine Ramp Consistent with RPEs**

(A) Delayed-reward task with odor cues. A subset of the animals shown in Figure 5 was used.

(B) Average firing rates (top, n = 67 neurons) and predicted calcium responses (bottom).

(C) The standard conditions in the linear track tasks for the same neurons as in (B) (n = 174 sessions from standard and experiments 1 and 3 from n = 67 neurons).

(D) Ramping slopes in the delayed-reward task (odor D, slope = –0.06 ± 0.02 spike/s$^2$, p = 0.018, n = 67) and the linear track task (average across tasks, slope = 0.10 ± 0.03 spike/s$^2$, p = 0.002, n = 67 neurons).

(E) Spatial cue manipulations (experiment 8) Left: patterned, distinct wall patterns were removed. Right: calcium signals in the standard (black, n = 6 mice) and patterned scene (orange). Arrows, medians of scene onsets.

*(legend continued on next page)*

significant test *R*). These results suggest that the concentration of dopamine still represents RPE.

### Cues that Indicate Reward Proximity Can Cause a Dopamine Ramp Consistent with RPEs

In previous experiments, ramping activity was not observed in delayed-reward tasks in which a reward was delivered with a relatively short fixed delay after a cue (e.g., Cohen et al., 2012; Starkweather et al., 2017). For more direct comparisons, we used a delayed-reward task using odor cues with delays that covered the timescale of the above tasks (Figure 7A; STAR Methods). A subset of the animals used in single-neuron recordings (Figure 5) also performed in the delayed-reward task in the same session while spikes were recorded continuously. Although anticipatory licking was similar (Figures S7A and S7B), the average spiking activity showed a slight negative ramp in the odor task and a positive ramp in the linear track task (Figures 7B–7D), demonstrating a task-specific modulation of dopamine neuron activity.

Some animals displayed locomotor activity during the linear track tasks (Figure 1G, bottom). However, locomotor activity and the magnitude of ramping were not correlated (Figure S2J; Figure 2L, right). Previous studies (Hamid et al., 2019; Mohebi et al., 2019) have suggested that dopamine ramps are present in an operant task (when the animal has to take a specific action to obtain a reward) but not in a Pavlovian task (in which a reward is delivered regardless of the animal's action). Technically, our linear track task is Pavlovian in design. However, it remains unclear how mice understand the task. We therefore designed additional experiments to clarify the task conditions that generate a dopamine ramp.

We trained a new set of mice in an operant contingency so that mice had to locomote a certain distance to obtain a reward. We observed a dopamine ramp similar to the Pavlovian design used above (Figure S7C). The results with teleportation (Figures S7D and S7H) and speed manipulations (Figures S7E and S7I) were also similar, indicating that the ramping activity in the operant context also represented RPEs. To assess the role of cues, we introduced other tracks in which animals could not estimate their position based on the visual input. The modified tracks had a uniform wall pattern (Figure 7E) or no pattern (Figure 7F). Although the animals showed anticipatory licking (Figure S7J), the timing of licking became broader (Figures S7F and S7G, top). Under these two manipulation conditions, we did not observe a positive ramp, although scene movement onset caused a small increase in calcium that somewhat persisted before reward (Figures 7F and 7G, bottom). Thus, whether the animal has to perform a certain movement to obtain a reward is not important for generating a dopamine ramp. Instead, the results highlight the importance of cues for the proximity to the reward.

To test whether a dopamine ramp could be generated by a non-navigational cue, we used a black horizontal bar that moved downward from the top of computer monitors. The animal received a drop of water when the bar reached a certain location (Figure 7H; Video S5). Consistent with our hypothesis, when the animal learned the task (Figure S7K), we observed a dopamine ramp similar to that in the linear track task (Figure S7L).

We next performed a series of experiments analogous to experiments 1–3 using moving bars. The results were largely consistent with those obtained using navigational stimuli (Figures 7I–7K; Figures S7M–S7O). Model fit analysis confirmed that the RPE model explained the data better than the value model (Figures 7L–7N). Furthermore, dopamine signals in the VS, measured using $GRAB_{DA2m}$, showed similar results (Figures 7O–7Q). These results demonstrate that dynamic cues that indicate proximity to reward can cause ramping dopamine neuron activity and that the ramping activity encodes the TD RPE.

## DISCUSSION

In this study, we developed a set of experimental paradigms to dissociate RPE and value, which are otherwise difficult to disambiguate. We found that teleportation and speed manipulations caused a phasic dopamine response and a change in the magnitudes of ramps, respectively, that were consistent with RPE. Furthermore, this was true at all of the stages of dopamine transmission we examined: spiking activity and calcium signals at somata, calcium signals at axons, and dopamine concentrations in the VS. We also showed that cues that indicate a gradual increase in temporal proximity to reward play an essential role in causing a dopamine ramp, regardless of whether the cue is

(F) Left: Solid-colored, optic flow was removed by a solid-colored stimulus (right). Shown are calcium signals in the standard (black, n = 6 mice) and solid-colored scene (orange).
(G) Ramping *R*s (p = 0.063 and 0.031 for experiments 8a and 8b, respectively; n = 6).
(H) Moving-bar experiment. A reward was delivered when the bar reached a target position (dotted line, for illustration purposes only).
(I) Experiment 1 with a moving bar. Calcium signals are aligned by long bar teleportation (red), short bar teleportation (orange), and pause (yellow) (n = 9 mice). Calcium signals under the standard condition (black) are aligned by reward onset (black dash).
(J) Experiment 2 (n = 10 mice). Vertical lines, teleportation.
(K) Experiment 3 (n = 11 mice).
(L) Comparisons of AIC between value and RPE models based on the exponential value function.
(M) Difference between the two models. All median ΔAICs, except for those for the standard condition, are significantly different from zero (p = [0.20 0.008 0.02 0.04 0.004], n = [9 8 8 9 9]).
(N) AICs from variants of models. Format as in Figure 3E.
(O–Q) Average dopamine signals measured using a dopamine sensor. Shown are (O) experiment 1 (n = 8 mice), (P) experiment 2 (n = 8 mice), and (Q) experiment 3 (n = 8 mice) using a moving bar.
See also Figure S7.

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013



navigational or more abstract or the task is operant or Pavlovian. We demonstrated that slowly fluctuating dopamine signals at a second timescale encode RPEs, similar to phasic dopamine signals. Our results thus provide a unified account of dopamine signals across these timescales as RPEs. Computing moment-by-moment changes in values is a hallmark of TD RPEs, which provides a solution to credit assignment problems in machine learning (Sutton and Barto, 1998). However, whether dopamine neurons in the brain compute RPE in this manner has not been tested. Our results demonstrate that dopamine neurons compute a derivative-like signal over value on a moment-by-moment basis, instantiating the central tenet of TD error signals.

### Experimental Design, Interpretation, and Limitations

The present study was originally motivated by the prediction that TD error can ramp up when the value function is sufficiently convex (Gershman, 2014). However, our interpretation is not bound to this hypothesis. Rather, the question is whether dopamine signals can be explained by a monotonically increasing function defined over space or by its derivative. More generally, we sought to empirically derive a parsimonious mathematical description of dopamine signals based on a set of parameters in the experiment. Our model-fitting results demonstrate that, by assuming the presence of a function that increases with proximity to reward, the dopamine signals can be described by its first derivative. Other observations, such as the dependency on reward amounts and the lack of response to sensory surprises, indicate that these functions match well with what we conceive as value and TD error. It is through these analyses that we inferred the potential value function under the "minimum" assumptions (the presence of a monotonically increasing function and allowing the dopaminergic activity to represent the [fractional] derivative of it). However, these analyses are not completely "assumption free." It will be interesting to address whether we can derive a set of mathematical equations for dopamine signals in a more data-driven manner from a more diverse set of potential functions (Schmidt and Lipson, 2009; Brunton et al., 2016).

It remains to be determined to what extent our results can be generalized. A recent study suggested that dopamine ramps can occur even without an explicit sensory cue when the reward is given after running a certain distance in a running wheel (Guru et al., 2020). It may be difficult to eliminate all sensory cues in such an experiment (e.g., the rotation of the wheel may provide a cue). Furthermore, it is possible that an animal's own movement acts as a cue (Killeen and Fetterman, 1988). In our experiment, a dopamine ramp was not observed without a positional cue even when the reward was given after running a fixed distance (Figures 7E–7G). However, we cannot exclude the possibility that, after sufficient training, a dopamine ramp may occur. Furthermore, it has been suggested that dopamine ramps become smaller after extensive training under some conditions (Guru et al., 2020). Our experiments typically contained test trials that might have prevented the animals from completely learning the standard condition. Furthermore, virtual reality does not completely reproduce a real environment. For instance, the spatial cues in our experiment may be less intuitive compared with those in a real environment, which might have reduced the effectiveness of learning. Finally, our results are consistent with a value function that is tied to position. This is likely because position, more so than other parameters in our experiment, served as a reasonable predictor of proximity to reward. If a different variable were a better predictor of proximity to reward, it is possible that the value function may, in that case, be a function of that variable rather than position, as suggested from our moving-bar experiment.

### The Origin of Ramping Dopamine Signals and the Potential Role of State Uncertainty

Although our results show that ramping dopamine signals are present in spiking activity of dopamine neurons, we cannot exclude the possibility that axonal modulation also has a role in shaping the signal. Dopamine release can be modulated locally at axons of dopamine neurons (Cachope et al., 2012; Threlfell et al., 2012; Zhou et al., 2001). It is generally difficult to make a direct comparison between somatic activity and neurotransmitter release *in vivo* because drawing a conclusion requires a complete match between the cell bodies and the axons that are observed and because there may be some nonlinear relationship between spiking activity and transmitter release.

However, ramping signals were present at the somata at the level of action potentials and associated calcium signals. This was also the case in dopamine neurons labeled retrogradely from the VS, where the axonal activity was monitored. Interestingly, we observed an ML gradient of ramping signals in the VTA (Figure 5D). The fact that one study (Mohebi et al., 2019) did not observe a ramp in dopamine neuron spiking may be due to the fact that the lateral region of the VTA was targeted in that study, underscoring the importance of recording from a matching pair of somata and axons.

Our study suggests that cues that indicate a gradual increase in proximity to reward play a key role in generating ramping signals. What then explains the difference between tasks in which dopamine signals ramp up or not? In reinforcement learning, values are computed based on the current "state" of the world. In this framework, the state is collectively defined by various types of information, such as the location, the objects that are present there, and the elapsed time from salient events. In natural situations, the state is often ambiguous because of partial information the animal receives, and state uncertainty can alter the shape of value functions (Babayan et al., 2018; Gershman and Uchida, 2019; Ludvig et al., 2008; Starkweather et al., 2017). The tasks we used in the present study are different from the delayed-reward tasks in terms of the structure of state uncertainty. In our virtual reality and moving-bar tasks, sensory inputs continuously provide information about the proximity to reward, indicating that the uncertainty about the current location (state) may stay constant or even decrease over time. In contrast, it is known that uncertainty about the proximity to reward increases proportionally with elapsed time without cues, following Weber's law (Gibbon et al., 1997). This difference in the structure of state uncertainty may result in different shapes of value functions between the tasks. It will be interesting to directly test these ideas theoretically (Mikhael et al., 2019) and experimentally in the future.

### Hierarchy in the Diversity of Dopamine Signals

Together with other evidence (Watabe-Uchida et al., 2017; Parker et al., 2016; Howe and Dombeck, 2016; Menegas et al., 2017, 2018), the present results reinforce the idea that dopamine signals in the VS resemble TD RPEs. Although dopamine neurons in this population exhibit some diversity (Engelhard et al., 2019; Kremer et al., 2020), at least some of the observed variability can be understood within the framework of RPEs. In contrast, more globally, a much larger level of diversity exists across different regions of the striatum (Cox and Witten, 2019;Watabe-Uchida and Uchida, 2018). Thus, the diversity of dopamine neurons is organized hierarchically. Looking at the diversity of dopamine signals in this hierarchical manner, it will be important to clarify whether relatively minor variations in dopamine signals within each subsystem have any functional significance. A recent study of distributional reinforcement learning indicates that relatively minor variability in RPE enables computational advantages (Dabney et al., 2020). Furthermore, it will be important to understand the computational principle by which each dopamine subsystem operates at a global level. Our experimental paradigms will be a powerful means by which to examine the nature of dopamine signals. A deeper understanding of the nature of dopamine signals will aid in addressing these questions in the future.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
- METHOD DETAILS
  - Surgery and virus injections
  - Virtual reality setup
  - Virtual linear track experiments in open-loop setting
  - Virtual linear track experiments in closed-loop setting
  - Delayed-reward task with odor cues
  - Moving-bar task
  - Number of animals and training history
- SUMMARY OF EXPERIMENTAL ANIMALS
  - Fiber fluorometry (photometry)
  - Electrophysiology
  - Histology
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Theoretical backgrounds
  - Quantitative analyses
  - Model fitting
  - Note on the shape of the fitted value function

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cell.2020.11.013.

## ACKNOWLEDGMENTS

We thank C. Harvey and D. Aronov for the virtual reality system. We thank B. Babayan, W. Menegas, and E. Soucy for fiber fluorometry setups. We thank A. Shakeel for histology. We thank J. Assad and K. Svoboda for comments. We also thank the members of the Uchida lab for discussions and critical reading of the manuscript. Some analyses were run on the FASRC Cannon cluster, supported by the FAS Division of Science Research Computing Group at Harvard University. This work was supported by NIH grants U19 NS113201 (to N.U. and S.J.G.), R01MH095953 (to N.U.), R01MH101207 (to N.U.), NS108740 (to N.U.), T32GM007753 (to J.G.M.), T32MH020017 (to J.G.M.), and U01NS103558 (to Y.L.); the Simons Collaboration on Global Brain (to N.U.); a Harvard Mind Brain and Behavior faculty grant (to S.J.G. and N.U.), a research fellowship from the Alfred P. Sloan Foundation (to S.J.G.); and the Junior Thousand Talents Program of China (to Y.L.).

## AUTHOR CONTRIBUTIONS

H.R.K. and N.U. conceived and supervised the entire project. H.R.K. conducted fiber fluorometry experiments and was involved in all other experiments. A.N.M. conducted electrophysiology and rabies fiber fluorometry. The electrophysiology experiment was designed by H.R.K., A.N.M., Paul Masset, and N.U. The electrophysiology data were analyzed by H.R.K. following data preparation and initial analysis by A.N.M. P.B. conducted some fiber fluorometry experiments. The dopamine sensor was developed by F.S., Y.Z., and Y.L. and tested by I.T.-K. and M.W.-U. J.G.M. and S.J.G. provided theoretical analysis and advised data analysis. H.R.K. analyzed the data and wrote the first draft. All the other authors discussed the results and commented on the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 15, 2019
Revised: August 20, 2020
Accepted: November 9, 2020
Published: November 27, 2020

## REFERENCES

Akaike, H. (1973). Information Theory as an Extension of the Maximum Likelihood Principle. In Second International Symposium on Information Theory, B.N. Petrov and F. Csaki, eds. (Akadémiai Kiadó), pp. 267–281.

Aronov, D., and Tank, D.W. (2014). Engagement of neural circuits underlying 2D spatial navigation in a rodent virtual reality system. Neuron *84*, 442–456.

Babayan, B.M., Uchida, N., and Gershman, S.J. (2018). Belief state representation in the dopamine system. Nat. Commun. *9*, 1891.

Bäckman, C.M., Malik, N., Zhang, Y., Shan, L., Grinberg, A., Hoffer, B.J., Westphal, H., and Tomac, A.C. (2006). Characterization of a mouse strain expressing Cre recombinase from the 3′ untranslated region of the dopamine transporter locus. Genesis *44*, 383–390.

Bayer, H.M., and Glimcher, P.W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron *47*, 129–141.

Bellman, R. (1954). The theory of dynamic programming. Bull. Am. Math. Soc. *60*, 503–515.

Berke, J.D. (2018). What does dopamine mean? Nat. Neurosci. *21*, 787–793.

Bromberg-Martin, E.S., Matsumoto, M., and Hikosaka, O. (2010). Distinct tonic and phasic anticipatory activity in lateral habenula and dopamine neurons. Neuron *67*, 144–155.

Brunton, S.L., Proctor, J.L., and Kutz, J.N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc. Natl. Acad. Sci. USA *113*, 3932–3937.

Cachope, R., Mateo, Y., Mathur, B.N., Irving, J., Wang, H.-L., Morales, M., Lovinger, D.M., and Cheer, J.F. (2012). Selective activation of cholinergic

interneurons enhances accumbal phasic dopamine release: setting the tone for reward processing. Cell Rep. *2*, 33–41.

Chen, G., King, J.A., Burgess, N., and O'Keefe, J. (2013a). How vision and movement combine in the hippocampal place code. Proc. Natl. Acad. Sci. USA *110*, 378–383.

Chen, T.-W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V., et al. (2013b). Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature *499*, 295–300.

Clark, J.J., Hollon, N.G., and Phillips, P.E.M. (2012). Pavlovian valuation systems in learning and decision making. Curr. Opin. Neurobiol. *22*, 1054–1061.

Cohen, J.Y., Haesler, S., Vong, L., Lowell, B.B., and Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature *482*, 85–88.

Cox, J., and Witten, I.B. (2019). Striatal circuits for reward learning and decision-making. Nat. Rev. Neurosci. *20*, 482–494.

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.

Dombeck, D.A., Khabbaz, A.N., Collman, F., Adelman, T.L., and Tank, D.W. (2007). Imaging large-scale neural activity with cellular resolution in awake, mobile mice. Neuron *56*, 43–57.

Engelhard, B., Finkelstein, J., Cox, J., Fleming, W., Jang, H.J., Ornelas, S., Koay, S.A., Thiberge, S.Y., Daw, N.D., Tank, D.W., and Witten, I.B. (2019). Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. Nature *570*, 509–513.

Eshel, N., Bukwich, M., Rao, V., Hemmelder, V., Tian, J., and Uchida, N. (2015). Arithmetic and local circuitry underlying dopamine prediction errors. Nature *525*, 243–246.

Farassat, N., Costa, K.M., Stojanovic, S., Albert, S., Kovacheva, L., Shin, J., Egger, R., Somayaji, M., Duvarci, S., Schneider, G., and Roeper, J. (2019). In vivo functional diversity of midbrain dopamine neurons within identified axonal projections. eLife *8*, e48408.

Fiorillo, C.D., Newsome, W.T., and Schultz, W. (2008). The temporal precision of reward prediction in dopamine neurons. Nat. Neurosci. *11*, 966–973.

Franklin, K.B., and Paxinos, G. (2008). The mouse brain in stereotaxic coordinates (Elsevier Academic Press).

Gershman, S.J. (2014). Dopamine ramps are a consequence of reward prediction errors. Neural Comput. *26*, 467–471.

Gershman, S.J., and Uchida, N. (2019). Believing in dopamine. Nat. Rev. Neurosci. *20*, 703–714.

Gibbon, J., Malapani, C., Dale, C.L., and Gallistel, C. (1997). Toward a neurobiology of temporal cognition: advances and challenges. Curr. Opin. Neurobiol. *7*, 170–184.

Guru, A., Seo, C., Post, R.J., Kullakanda, D.S., Schaffer, J.A., and Warden, M.R. (2020). Ramping activity in midbrain dopamine neurons signifies the use of a cognitive map. bioRxiv. https://doi.org/10.1101/2020.05.21.108886.

Hamid, A.A., Pettibone, J.R., Mabrouk, O.S., Hetrick, V.L., Schmidt, R., Vander Weele, C.M., Kennedy, R.T., Aragona, B.J., and Berke, J.D. (2016). Mesolimbic dopamine signals the value of work. Nat. Neurosci. *19*, 117–126.

Hamid, A.A., Frank, M.J., and Moore, C.I. (2019). Dopamine waves as a mechanism for spatiotemporal credit assignment. bioRxiv. https://doi.org/10.1101/729640.

Harvey, C.D., Collman, F., Dombeck, D.A., and Tank, D.W. (2009). Intracellular dynamics of hippocampal place cells during virtual navigation. Nature *461*, 941–946.

Howe, M.W., and Dombeck, D.A. (2016). Rapid signalling in distinct dopaminergic axons during locomotion and reward. Nature *535*, 505–510.

Howe, M.W., Tierney, P.L., Sandberg, S.G., Phillips, P.E.M., and Graybiel, A.M. (2013). Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. Nature *500*, 575–579.

Ikemoto, S. (2007). Dopamine reward circuitry: two projection systems from the ventral midbrain to the nucleus accumbens-olfactory tubercle complex. Brain Res. Brain Res. Rev. *56*, 27–78.

Killeen, P.R., and Fetterman, J.G. (1988). A behavioral theory of timing. Psychol. Rev. *95*, 274–295.

Kim, H.R., Angelaki, D.E., and DeAngelis, G.C. (2015). A functional link between MT neurons and depth perception based on motion parallax. J. Neurosci. *35*, 2766–2777.

Kobayashi, S., and Schultz, W. (2008). Influence of reward delays on responses of dopamine neurons. J. Neurosci. *28*, 7837–7846.

Kremer, Y., Flakowski, J., Rohner, C., and Lüscher, C. (2020). Context-Dependent Multiplexing by Individual VTA Dopamine Neurons. J. Neurosci. *40*, 7489–7509.

Kvitsiani, D., Ranade, S., Hangya, B., Taniguchi, H., Huang, J.Z., and Kepecs, A. (2013). Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. Nature *498*, 363–366.

Lammel, S., Hetzel, A., Häckel, O., Jones, I., Liss, B., and Roeper, J. (2008). Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. Neuron *57*, 760–773.

Lima, S.Q., Hromádka, T., Znamenskiy, P., and Zador, A.M. (2009). PINP: a new method of tagging neuronal populations for identification during in vivo electrophysiological recording. PLoS ONE *4*, e6099.

Lloyd, K., and Dayan, P. (2015). Tamping Ramping: Algorithmic, Implementational, and Computational Explanations of Phasic Dopamine Signals in the Accumbens. PLoS Comput. Biol. *11*, e1004622.

Ludvig, E.A., Sutton, R.S., and Kehoe, E.J. (2008). Stimulus representation and the timing of reward-prediction errors in models of the dopamine system. Neural Comput. *20*, 3034–3054.

Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat. Neurosci. *13*, 133–140.

Menegas, W., Babayan, B.M., Uchida, N., and Watabe-Uchida, M. (2017). Opposite initialization to novel cues in dopamine signaling in ventral and posterior striatum in mice. eLife *6*, e21886.

Menegas, W., Akiti, K., Amo, R., Uchida, N., and Watabe-Uchida, M. (2018). Dopamine neurons projecting to the posterior striatum reinforce avoidance of threatening stimuli. Nat. Neurosci. *21*, 1421–1430.

Mikhael, J.G., Kim, H.R., Uchida, N., and Gershman, S.J. (2019). Ramping and state uncertainty in the dopamine signal. bioRxiv. https://doi.org/10.1101/805366.

Mohebi, A., Pettibone, J.R., Hamid, A.A., Wong, J.T., Vinson, L.T., Patriarchi, T., Tian, L., Kennedy, R.T., and Berke, J.D. (2019). Dissociable dopamine dynamics for learning and motivation. Nature *570*, 65–70.

Morita, K., and Kato, A. (2014). Striatal dopamine ramping may indicate flexible reinforcement learning with forgetting in the cortico-basal ganglia circuits. Front. Neural Circuits *8*, 36.

Niv, Y. (2009). Reinforcement learning in the brain. J. Math. Psychol. *53*, 139–154.

Parker, N.F., Cameron, C.M., Taliaferro, J.P., Lee, J., Choi, J.Y., Davidson, T.J., Daw, N.D., and Witten, I.B. (2016). Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. Nat. Neurosci. *19*, 845–854.

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 *422*, 614–618.

Podlubny, I. (1998). Fractional differential equations: an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications (Academic Press).

Reardon, T.R., Murray, A.J., Turi, G.F., Wirblich, C., Croce, K.R., Schnell, M.J., Jessell, T.M., and Losonczy, A. (2016). Rabies Virus CVS-N2c(ΔG) Strain

Enhances Retrograde Synaptic Transfer and Neuronal Viability. Neuron *89*, 711–724.

Roitman, M.F., Stuber, G.D., Phillips, P.E.M., Wightman, R.M., and Carelli, R.M. (2004). Dopamine operates as a subsecond modulator of food seeking. J. Neurosci. *24*, 1265–1271.

Schmidt, M., and Lipson, H. (2009). Distilling free-form natural laws from experimental data. Science *324*, 81–85.

Schmitzer-Torbert, N., and Redish, A.D. (2004). Neuronal activity in the rodent dorsal striatum in sequential navigation: separation of spatial and reward responses on the multiple T task. J. Neurophysiol. *91*, 2259–2272.

Schultz, W. (2019). Recent advances in understanding the role of phasic dopamine activity. F1000Res. *8*, 1680.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.

Stalnaker, T.A., Howard, J.D., Takahashi, Y.K., Gershman, S.J., Kahnt, T., and Schoenbaum, G. (2019). Dopamine neuron ensembles signal the content of sensory prediction errors. eLife *8*, e49315.

Starkweather, C.K., Babayan, B.M., Uchida, N., and Gershman, S.J. (2017). Dopamine reward prediction errors reflect hidden-state inference across time. Nat. Neurosci. *20*, 581–589.

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.e19.

Sun, F., Zhou, J., Dai, B., Qian, T., Zeng, J., Li, X., Zhuo, Y., Zhang, Y., Wang, Y., Qian, C., et al. (2020). Next-generation GRAB sensors for monitoring dopaminergic activity in vivo. Nat. Methods *17*, 1156–1166.

Sutton, R.S. (1988). Learning to predict by the methods of temporal differences. Mach. Learn. *3*, 9–44.

Sutton, R.S., and Barto, A.G. (1998). Reinforcement Learning: An introduction (MIT Press).

Takahashi, Y.K., Batchelor, H.M., Liu, B., Khanna, A., Morales, M., and Schoenbaum, G. (2017). Dopamine Neurons Respond to Errors in the Prediction of Sensory Features of Expected Rewards. Neuron *95*, 1395–1405.e3.

Threlfell, S., Lalic, T., Platt, N.J., Jennings, K.A., Deisseroth, K., and Cragg, S.J. (2012). Striatal dopamine release is triggered by synchronized activity in cholinergic interneurons. Neuron *75*, 58–64.

Tsai, H.-C., Zhang, F., Adamantidis, A., Stuber, G.D., Bonci, A., de Lecea, L., and Deisseroth, K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. Science *324*, 1080–1084.

Uchida, N., and Mainen, Z.F. (2003). Speed and accuracy of olfactory discrimination in the rat. Nat. Neurosci. *6*, 1224–1229.

Watabe-Uchida, M., and Uchida, N. (2018). Multiple Dopamine Systems: Weal and Woe of Dopamine. Cold Spring Harb. Symp. Quant. Biol. *83*, 83–95.

Watabe-Uchida, M., Eshel, N., and Uchida, N. (2017). Neural Circuitry of Reward Prediction Error. Annu. Rev. Neurosci. *40*, 373–394.

Zhou, F.M., Liang, Y., and Dani, J.A. (2001). Endogenous nicotinic cholinergic activity regulates dopamine release in the striatum. Nat. Neurosci. *4*, 1224–1229.

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013



# STAR★METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Bacterial and Virus Strains | | |
| AAV5-CAG-FLEX-jGCaMP6m-WPRE-SV40 | UPenn vector core | N/A |
| AAV9-hSyn-DA2m | Vigene Biosciences | N/A |
| AAV5-EF1a-DIO-hChR2(H134R)-EYFP | UNC Vector Core | N/A |
| Experimental Models: Organisms/Strains | | |
| Mouse: B6.SJL-Slc6a3$^{tm1.1(cre)Bkmn}$/J | The Jackson Laboratory | Jax #006660; RRID:IMSR_JAX:006660 |
| Mouse: B6.Cg-Gt(ROSA)$^{26Sortm9(CAG-tdTomato)Hze}$/J | The Jackson Laboratory | Jax #007909; RRID:IMSR_JAX:007909 |
| Mouse: .C57BL/6 J | The Jackson Laboratory | Jax # 000664; RRID:IMSR_JAX:000664 |
| Software and Algorithms | | |
| VirMEn (Virtual Reality MATLAB Engine) | Dmitriy Aronov | https://pni.princeton.edu/pni-software-tools/virmen |
| MATLAB (2018b for analysis, 2011b for VirMEn) | MathWorks | https://www.mathworks.com/ |
| MClust software (version 4.3.02) | A. David Redish | http://redishlab.neuroscience.umn.edu/MClust/MClust.html |
| Open Ephys GUI | Open Ephys | https://open-ephys.org/gui |
| Other | | |
| Isosol (Isoflurane, USP) | Vedco | N/A |
| Ketoprofen (for analgesia) | Patterson Veterinary | Cat #07-803-7389 |
| Buprenorphine | Patterson Veterinary | Cat #07-850-2280 |
| LRS-0473 DPSS Laser System | LaserGlow Technologies | Cat #R471003FX |
| Sandvik Kanthal HP Reid Precision Fine Tetrode Wire | Sandvik | Cat #PF000591 |
| Mono Fiber-optic Cannulas | Doric Lenses | MFC_200/245-0.53_5mm_MF1.25_FLT |
| 32channel amplifier board | Intan Technologies | RHD2132 |
| Open Ephys Acquisition Board | Open Ephys | https://open-ephys.org/acq-board |

## RESOURCE AVAILABILITY

### Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Hyung-Goo Kim (HyungGoo.R.Kim@gmail.com).

### Materials Availability

This study did not generate new unique reagents.

### Data and Code Availability

Data and analysis code will be available in some form for purpose of reproducing or extending the analysis. MATLAB codes for data visualizations can be found at https://github.com/hkim09/libkm.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Mice

A total of 69 adult mice were used in the experiments. Thirty-six adult male mice were used for the experiments using calcium indicator (GCaMP). Twenty-nine mice were heterozygous for the gene expressing the Cre recombinase under the control of the promoter of the dopamine transporter (DAT) gene (Bäckman et al., 2006) (DAT-Cre or B6.SJL-Slc6a3$^{tm1.1(cre)Bkmn}$/J mice; The Jackson Laboratory; RRID:IMSR_JAX:006660). Seven mice were the result of a cross between DAT-Cre mice and a tdTomato transgenic line such

that they were heterozygous for DAT-Cre and also heterozygous for tdTomato (Gt(ROSA)$^{26Sortm9(CAGtdTomato)Hze}$, Jackson Laboratory) (Madisen et al., 2010). We did not observe a difference in results between these mice, thus results were combined. Six adult C57/BL6J DAT-Cre male mice were used for GFP control experiments. Ten adult C57/BL6J wild-type male mice were used for the experiments using the dopamine sensor. Four adult C57/BL6J DAT-Cre male mice were used for the experiments using rabies expressing GCaMP. Thirteen adult C57/BL6J DAT-Cre male mice were used for electrophysiological recording experiments. All mice were backcrossed for > 5 generations with C57/BL6J mice. Animals were singly housed on a 12 hr dark/12 hr light cycle (dark from 07:00 to 19:00). All procedures were performed in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and approved by the Harvard Animal Care and Use Committee.

## METHOD DETAILS

### Surgery and virus injections

#### *Surgery for fiber fluorometry of GCaMP signals in the VS*

To prepare animals for recording, we performed a single surgery with three key components: (1) injection of an adeno associated virus (AAV) that expresses GCaMP in the presence of Cre recombinase (AAV-FLEX-GCaMP) into the midbrain, (2) head-plate installation, and (3) implantation of one or more optical fibers into the striatum (Babayan et al., 2018; Menegas et al., 2017). At the time of surgery, all mice were 2–4 months old. All surgeries were performed under aseptic conditions with animals anesthetized with isoflurane (1- 2% at 0.5 - 1.0 L/min). Analgesia (ketoprofen for post-surgery treatment, 5 mg/kg, I.P.; buprenorphine for pre-operative treatment, 0.1 mg/kg, I.P.) was administered for 3 days following each surgery. We removed the skin above the surface of the brain and dried the skull using air. To express GCaMP specifically in dopamine neurons, we unilaterally injected 250 nL of AAV5-CAG-FLEX-GCaMP6m ($1 \times 10^{12}$ particles/ml, Penn Vector Core) into both the VTA and SNc (500 nL total). To target the VTA, we made a small craniotomy and injected virus at bregma 3.1, lateral 0.6, depths 4.4 and 4.1 mm. To target SNc, we injected virus at bregma 3.3, lateral 1.6, depths 3.8 and 3.6 mm. Virus injection lasted several minutes, and then the injection pipette was slowly removed over the course of several minutes.

We then installed a head-plate for head-fixation by gluing a head-plate onto the top of the skull (C&B Metabond, Parkell). We used ring-shaped head-plates to ensure that the skull above the striatum would be accessible for fiber implants. Finally, during the same surgery, we also implanted optical fibers into the VS (nucleus accumbens core). For a subset of animals, we also implanted fibers into dorsomedial or dorsolateral striatum. To do this, we first slowly lowered optical fibers (200 μm diameter, Doric Lenses) into the striatum using a fiber holder (SCH_1.25, Doric Lenses). The coordinates we used for targeting were bregma 1.0, lateral 1.1, depth 4.1 mm. Once fibers were lowered, we first attached them to the skull with UV-curing epoxy (Thorlabs, NOA81), and then a layer of black Ortho-Jet dental adhesive (Lang Dental, IL). After waiting for fifteen minutes for this glue to dry, we applied a small amount of rapid-curing epoxy (A00254, Devcon) to attach the fiber cannulas to the underlying glue and head-plate. After waiting for fifteen minutes for the epoxy to cure, the surgery was completed.

#### *Surgery for fiber fluorometry of GCaMP signals in the VTA*

The above results suggest that the somewhat larger ramping of axonal calcium signals may be due to the slow kinetics of calcium signals compared to spiking activities. Furthermore, the ramping in spiking activity at the cell body suggests that ramping axonal calcium activities in the VS may originate from ramping somatic spiking. If these were true, then calcium signals measured at the cell bodies of dopamine neurons should also exhibit a similar level of ramping as those at the axons, as well as the features of RPEs that we observed in the above experiments. For that purpose, we recorded calcium signals from dopamine neurons in the VTA. The injection of virus was done in the same way as the recordings in the VS. For recording in the VTA, we placed a fiber at bregma 3.1, lateral 0.6, depths 4.1 mm.

#### *Surgery for rabies anatomical tracing and GCaMP fluorometry*

Dopamine neurons projecting to the nucleus accumbens core were retrogradely labeled by a modified rabies virus that expresses a calcium indicator (Reardon et al., 2016). Adult male DAT-Cre mice were first injected with 0.5μL AAV5-FLEX-TVA-mCherry (Custom AAV preparation by UNC vector core) to the left medial VTA (AP −3.05, ML 0.4, depth 4.3-4.5 mm). Two weeks later, mice were injected with 0.5μL of a 1:2 mixture of AAV1-CAG-BFP (Custom AAV preparation by UNC vector core) and CVS-N2c$^{\Delta G}$-GCaMP6f (EnvA) (provided by Catherine Dulac's laboratory, obtained from HHMI Janelia) to the left VS (AP +1.1, ML +1.0, depth 3.85-3.95 mm) and then an optical fiber was implanted within the left medial VTA (AP −3.05, ML 0.4, depth 4.35 mm).

#### *Surgery for fiber fluorometry of dopamine sensor signals*

To directly examine whether the dopamine concentration in the VS represents RPE or value, we performed direct measurements of dopamine using a genetically encoded dopamine sensor (a second-generation sensor, $GRAB_{DA2m}$) (Sun et al., 2018, 2020). This dopamine sensor ($GRAB_{DA2m}$) was expressed in neurons in the VS, and fluorescent signals were measured through an optical fiber implanted in the VS (Figure 6E). Surgical procedures up to virus injection were the same as GCaMP injections described above. Instead of GCaMP, we injected 400 nL of AAV9-hSyn-DA2m (Vigene bioscience) into the VS (bregma 1.0, lateral 1.1, depths 4.2 and 4.1mm). Once injection was completed, we implanted an optical fiber and head plate in the same way as the surgery for GCaMP recording.

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013



### Surgery for single unit recording of opto-tagged dopamine neurons

We performed two surgeries, both stereotactically targeting the left VTA (from bregma: 3.1 mm posterior, 0.6 mm lateral, 4.2 mm ventral). In the first surgery, we injected 500 nL of AAV5 carrying an inverted ChR2-encoding sequence (H134R) fused to the sequence expressing the fluorescent reporter eYFP and flanked by double *loxP* sites (AAV5-DIO-ChR2-EYFP) (Tsai et al., 2009). We previously showed that the expression of this virus is highly selective and efficient in dopamine neurons (Cohen et al., 2012). After 2 weeks, we performed the second surgery to implant a head plate and custom-built microdrive containing 8 tetrodes and an optical fiber.

### Virtual reality setup

Virtual environments were displayed on three liquid crystal display (LCD) monitors with thin frames (width 53 cm, height 30 cm) that were placed on the left, front, and right side of animals (Chen et al., 2013a; Harvey et al., 2009). A workstation computer with a high-performance graphics card (NVIDIA Quadro K2200) was used to present visual images. ViRMEn software (Aronov and Tank, 2014) was used to generate virtual objects and render visual images using perspective projection. No dropping of image frames was confirmed by measuring intervals in the inter-frame callback function, and by photodiode measurements while a test program alternated the brightness of the screen every frame.

Animals were head-restrained at the center of three monitors, 7.5 cm above the bottom of the screen. Mice were placed on a cylindrical styrofoam treadmill (diameter 20.3 cm, width 10.4 cm). The rotational velocity of the treadmill was encoded using a rotary encoder. The output pulses of the encoder were converted into continuous voltage signals using a customized Arduino program running on a microprocessor (Teensy 3.2).

Water reward was given through a water spout located in front of the animal's mouth. Licking tongue movements were monitored using an infrared sensor (OPB819Z, TT Electronics). Voltage signals from the rotary encoder and the lick sensor were digitized into a PCI-based data-acquisition system (PCIe-6323, National Instruments) installed on the visual stimulation computer. Timing and amount of water were controlled through a micro-solenoid valve (LHDA 1221111H, The Lee Company) and switch (2N7000, On Semiconductor). Analog output TTL pulse was generated from the visual stimulation computer to deliver reward to the animals.

### Virtual linear track experiments in open-loop setting

Animals were trained in a virtual linear track (Figure 1A; length of 150 arbitrary units, a.u., and width of 30 a.u.). The maze was composed of a starting platform and a corridor with walls on both sides. The walls have four different texture patterns to help animals recognize a position in the virtual space (Video S1).

We first trained animals on the standard approach-to-target task to learn the association between target location and reward. Once the animals learned the task, we ran a series of tasks with test trials to examine the nature of dopamine signals. We typically ran each task for two consecutive days (with zero- or one-day break). A daily session started with 5-10 standard trials to help animals remember the task before presenting any test trials. Unless otherwise noted, unexpected reward (5 μl) was given during the inter-trial interval on 3%–6% of trials.

#### Standard approach-to-target task

The session started in a dark gray background. Trials started with the presentation of the visual scene with the animal placed at the starting position (0 a.u.). After a random delay (1 s offset plus a random delay drawn from a modified exponential distribution with a mean of 1.5 s and cutoff of 3.5 s), the visual scene started moving forward. The velocity linearly increased for 1 s until it reached 13 a.u./s, after which the velocity was maintained to be constant until the animal reached the target position (97 a.u.). A drop of water reward (5 μl) was delivered through a water spout that was placed in front of the animal's mouth. No reward omission trials were used. Once the reward was given, the visual stimulus was turned off after a random delay (drawn from an exponential distribution with a mean of 1 s, to which a 1 s offset was added). Delay was re-drawn if it exceeded 4 s. Inter-trial interval (ITI) was drawn from an exponential distribution with a mean of 3 s, to which a 3.5 s offset was added. ITI was re-drawn if it exceeded 10 s.

#### Experiment 1 (Teleport to the same destination & pausing)

The following three types of test conditions were randomly interleaved with the standard condition. Each test condition comprised 20% of trials.

1. *Long teleport*: when the animal reached a predefined position (40 a.u.), it was teleported to a position closer to the reward (70 a.u.). At the time of teleport, the screens were briefly (93 ms) blanked to black.
2. *Short teleport*: when the animal reached another predefined position (65 a.u.), it was teleported to the same destination (70 a.u.).
3. *Pause*: when the animal reached this destination (70 a.u.), the progression of the scene was paused for 5 s, after which the scene movement resumed.

In all trials, the animal received a reward when it reached the same goal location (97 a.u.).

#### Experiment 2 (Three-teleport task)

The following three types of teleport conditions were randomly interleaved with the standard condition with the frequency of all teleport conditions being 33%–40% in total. On a fraction of trials, when the animal reached one of the three teleport positions (5, 25, and

45 a.u.), it was teleported by the same distance (30 a.u.) forward to the location closer to the goal location (35, 55 and 75 a.u., respectively). The screens were briefly (93 ms) blanked to black at the time of teleport. Scene progression was 20% slower than that it had been during training, except for the first four animals. Data from these four animals were excluded from the population time courses but included in other statistical analyses.

***Experiment 3 (Speed manipulation)***

On a fraction of trials (20%), the progression of the scene was twice as fast as in the standard condition. On some other trials (20%), the progression was half as fast as in the standard condition. The rest of the trials were the same as the standard approach-to-target task. The acceleration of the scene in the manipulated condition was identical to the standard condition (13 a.u./$s^2$).

*No Reward control experiments.* We tested whether the transient response to teleport is due to the short blackout or the abrupt change of visual scene in the absence of a reward context. Before training began with reward in the standard approach-to-target condition, a subset of animals performed Experiment 1 (n = 7 mice), Experiment 2 (n = 13 mice), and Experiment 3 (n = 13 mice) without reward at the target location. Unexpected rewards were delivered during the inter-trial intervals on 10% of trials. We ran one session for each protocol (70-100 trials), one or two control experiments a day.

***Experiment 4 (Dynamic speed manipulation)***

In the speed manipulation task (Experiment 3), animals might use the speed of scene movement as a cue for an early or late reward (i.e., time to reward). To exclude the possibility that the dopamine signals are driven by time to reward signaled by scene movement speeds. we dynamically changed the speed of scene within a trial while maintaining the total time between trial start and reward (Figure S3I) constant. We used three trial types: (1) accelerating, (2) decelerating, and (3) standard conditions. The instantaneous speeds followed the equations below (Kim et al., 2015):

$$v_1(t) = \sin\left(\frac{t + t_{off}}{2t_0} * 2 * pi + \theta_i\right)$$

$$v_2(t) = v_{amp}\left(|v_1(t)|^k sign(v_1(t)) + 1\right) + v_{off,i}$$

$$G(t) = \exp\left(-\left(\frac{(t - t_0)}{\sigma_{vel}}\right)^{n_{vel}}\right) v(t) = G(t) * v_2(t)$$

where $i = 1, 2$ for the two opposite phases, $t_0 = 6$, $t_{off} = 1.5$, $\sigma_{vel} = 5$, $n_{vel} = 12$, $v_{amp} = 10$, $v_{off,1} = 4.309$, $v_{off,2} = 4.641$, $\theta_1 = 0$, $\theta_2 = \pi$, $k = 0.7143$. The animals either decelerate (i = 1) or accelerate (i = 2) near the reward location while the trial duration is identical to the standard condition (Figure 2M).

***Experiment 5a (Between-track teleport)***

We tested whether the abrupt change of visual scene caused a transcient excitation. Once the animals were fully trained for the standard linear track (Track 1; Figure 4A, left), we introduced the second linear track (Track 2; Figure 4A, right) in which wall patterns were distinct from Track 1 (Video S4). In both tracks, animals obtained the same amount of water (4 μL) at the end of the tracks (97 a.u.). On 60% of the trials, mice performed the task in Track 1. On 3/8 of the trials in each track, mice were teleported to the same position in the other tack (Figure 4B; 55 a.u.).

***Experiment 5b (Between-track teleport & forward teleport)***

After collecting data in Experiment 5a, we used the same animals to confirm that a forward teleport elicited a transcient excitation in the same session (Figure 4E). Mice started trials in Track 1 on 50% of the trials and started trials in Track 2 on the rest of trials. On 1/3 of the trials initiated in Track 1, mice were teleported forward from 40 a.u. to 70 a.u.. On the other 1/3 of the trials in Track 1, mice were teleported from Track 1 to Track 2 (60 a.u.). Mice were neither teleported forward nor to Track 1 when trials were started in Track 2.

***Experiment 5c (Between-tracks teleport with different reward amounts)***

Animals obtained 2 μL in Track 1 and 12 μL in Track 2 (Figure 4H). On a faction of trials (1/6 for each condition), animals were teleported forward (40 a.u.), teleported from Track 1 to Track 2 (60 a.u.), or teleported from Track 2 to Track 1 (60 a.u.). Forward-teleport trials were not used for analyses.

***Experiment 6 (Forward-and-backward teleport)***

We randomly interleaved the standard, forward teleport (from 40 a.u. to 70 a.u.), and backward teleport (from 70 a.u. to 40 a.u.; Figure 4K) conditions. Each manipulated condition comprised 20% of trials.

***Experiment 7 (Reward-size manipulation)***

We alternated blocks of 25 trials to switch between a small-reward (2.25 μl/trial) and a large-reward (10 μl/trial) conditions (Figure 4N). No explicit cue was given to notify a block switch. On the first day of Experiment 7, we randomly chose whether the first block was a small- or large-reward block. The other reward size was used for the first block on the next day. On a fraction of trials (20%), we teleported the animals from 45 a.u. to 75 a.u..

**Virtual linear track experiments in closed-loop setting**

In the main experiments describe above (Experiments 1-7), the scene movement was decoupled from the animal's movement, and reward was delivered regardless of the animal's action ("open-loop" or Pavlovian). In this paradigm, the scene movement was coupled to the animal's locomotion and the animal was required to locomote a predefined distance to obtain reward ("closed

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013



loop'' or operant). The virtual environment was identical to Track 1, except that the length of the track was increased from 150 a.u. to 1000 a.u.. The modification was made to prevent the animals from using looming cues at the end of the track to estimate their positions.

***Training standard approach-to-target task***

After habituation in the experimental setup, mice were trained in a 'running exercise' protocol for 1-2 days. In this training, mice obtained 3 μL of water with an interval reversely proportional to their running speed, with the lower bound of 1.5 s. After mice achieved a consistent running behavior (mice are in locomotion state for more than 30% of the entire session and average running speed is greater than 5 cm/s), we introduced a closed-loop linear track in a staircase manner. At a trial start, mice were placed on one of the eight starting positions (0, 20, 30, 40, 50, 60, 70, 75 a.u.; the reward location, 97 a.u.). On the first day of training, trials were started with the starting position (75 a.u.) that is the closest to the reward location. The starting position of the next trial is probabilistically determined depending on the performance of the current trial: after a successful trial, the mouse was located one step farther from the reward location, with probability of 0.2. After an unsuccessful trial, the mouse was located one step closer to the reward location with the probability of 0.5. Otherwise, the starting location remained the same in the next trial. Trials were aborted and marked unsuccessful if either the animal did not leave the start location in 20 s or the animal did not reach the target position in 40 s after leaving the start location. From the second day of training, the start location of the first trial was set to be two steps closer than the start location with a stable performance in the previous session. Mice gradually learned to run consistently through this probabilistic staircase procedure.

Once the animal achieved stable running behavior at the starting position of 0 a.u., we ran an adaptive gain-adjustment protocol. This protocol adaptively changed the visual gain of indivual aniamls such that the interval between locomotion onset and reward delivery (locomotion_duration) became similar to that in the open loop condition (7.5 s). Each trial, we calculated the target gain:

$$\text{Target gain} = \text{locomotion\_duration}/7.5$$

and updated the gain adaptively with a learning rate ($\alpha_i$) that exponentially decreases over trials:

$$\text{gain}_{i+1} = \text{gain}_i + \alpha_i * (\text{target gain}_i - 1)$$

$$\alpha_i = 0.1 * \exp(-\textit{trial number}/70)$$

Once a converging adaptive gain was obtained in one or two sessions, the gain was fixed for further training and experiments.

Once the animals showed stable performance (success rate > 85%) and anticipatory licking, the mice were recorded in the standard experiments and the experiments with test trials (Experiments 1 and 3). After testing, some mice were further tested in Experiment 8 as described below.

***Experiments 1 and 3***

We conducted the teleport and pause experiment (Experiment 1) and the speed-manipulation experiment (Experiment 3) to examine the nature of dopamine signals in a closed-loop virtual reality environment. The experimental manipulations (e.g., the proportion of manipulated trials, the dynamics of position in teleport and pause) were identical to the open-loop experiments.

***Experiment 8 (Spatial cue manipulation)***

To examine the role of cues that indicate the proximity to reward in generating ramping dopamine signals, we manipulated the visual cues in the standard virtual reality task. We prepared two linear tracks that differed in the wall pattern compared to the original track (Track 1). Track 3 contained a regular pattern of black spots on blue walls (patterned; Figure 7E, left bottom; Experiment 8a), which eliminated cues associated with specific locations in the linear track, but created an optic flow. Track 4 consisted of solid-blue walls without the floor pattern, which generated no optic flow in the visual scene (solid-colored; Figure 7F, left bottom; Experiment 8b). Animals were required to run the same distance to obtain a reward on Tracks 1, 3 and 4. The test tracks (Track 3 or 4) were randomly interleaved (25% - 33%) with the standard track (Track 1) during training and recording.

## Delayed-reward task with odor cues

In previous experiments from our laboratory, ramping dopamine neuron activity was not observed in delayed-reward tasks in which reward was delivered with a fixed delay after a cue (e.g., Cohen et al., 2012; Starkweather et al., 2017). We sought to compare the activity of dopamine neurons in the present task to the activity of dopamine neurons in a task in which ramping dopamine neuron activity was not previously observed. One difference between the present and previous experiments is the timescale: whereas there was a relatively long time delay (~7 s) between scene movement onset and reward in the present study, the previous studies have used relatively short delays (e.g., 2-3 s) (but see Fiorillo et al., 2008; Kobayashi and Schultz, 2008). Once the mice were fully trained for the approach-to-target task, a subset of mice used for single neuron recordings were trained in a classical conditioning (a delayed-reward task) using odor cues (6 out of 13 mice). We used the same behavioral apparatus with three monitors as the virtual reality experiments, additionally equipped with a custom-made olfactometer (odor delivery machine) (Uchida and Mainen, 2003). A trial start was signaled by a brief (0.25 s) flash of green background on the computer monitors. 1.25 s after the onset of trial start cue, an odor was delivered for 0.5 s, followed by a delivery of water reward. Different odors indicated different delays between an odor onset and

reward (0.6 s, 1.5 s, 3.75 s, 9.375 s for Odors A, B, C and D, respectively). Only trials with Odor C and D are reported here to match the timescales. Inter-trial intervals were adjusted for each odor cue such that the intervals between trial start signals (flash of light) were similar across odor cues on average, and were between 17 and 20 s. Animals showed anticipatory licking after three to seven days of training. Once we found a light-identified neuron on mice that were fully trained for both the linear track task and the odor-reward association task, we first ran 70-90 trials of the odor-reward association task and then ran a series of standard and manipulation experiments in the linear track. For more direct comparisons to the virual reality tasks, we only show results using two odors (Odors C and D) which have similar delays as in the linear track tasks.

Odors were delivered similar to previous studies (Uchida and Mainen, 2003). Each odorant was dissolved in mineral oil at 1/10 dilution. 30 μl of odor solution was placed on a glass fiber filter paper. Filtered air was run through the filter paper to produce a total flow rate of 1 L/min. Odorants used included isoamyl acetate, (+)-carvone, 1-hexanol, p-cymene, ethyl butyrate, 1-butanol, limonene, dimethoxybenzene, caproic acid, 4-heptanone and eugenol. A set of odorants used for each mouse was assigned randomely.

### Moving-bar task

We used the same three-monitor display setup as in our linear track tasks. The background was gray. A black ring-shaped object was used to render black bars on the three screens in the virtual environment. The object (2.5 cm vertical thickness) moved vertically at a constant speed to indicate reward proximity (Video S5).

#### *Standard task*

A black bar was initially presented at the top of the screen. After a random inter-trial interval, the bar started to move from the top to the bottom of the screen at a constant speed (3.7 cm/s). When the bar reached a goal position (25 cm from the top of the screen, 6.7 s after the movement onset of the bar), a drop of water (5 μl) was delivered. After a random delay (the same as the 1D maze task), the bar was shifted back to the original starting position. The screen was kept on during the inter-trial interval.

#### *Experiment 1 with moving bar: Two teleport and one pause task*

On some fraction of trials (12.5%, respectively), when the position of the bar reached one of two positions on the screen (10.9 cm or 16.25 cm from the top of the screen), the bar was abruptly shifted downward by 6.25 cm or 0.93 cm, respectively, and maintained movement with constant speed. On another 12.5% of trials, the bar movement was paused for 5.0 s, after which the movement resumed. These test conditions were randomly interleaved with the standard condition.

#### *Experiment 2 with moving bar: Three-teleport task*

When the center of the bar reached 1.56, 6.9, or 12.18 cm from the top of the screen (0.4 s, 1.37 s, or 2.78 s from bar movement onset, respectively), the bar shifted its position by 6.6 cm (12.5% of total trials, respectively). The teleport conditions were randomly interleaved with the standard condition.

#### *Experiment 3 with moving bar: Speed manipulation task*

The movement of the bar was twice as fast as the standard condition (20%) or half as fast as the standard condition (20%).

### Number of animals and training history

The vertical order in the table below represents approximate order of experiments for each animal with the identity of each animal indicated by a unique number in the Animal column. The number of animals vary across experiments, and not all animals were subjected to all possible experiments, because (1) some experiments were developed later than others (e.g., Experiment 1 was established later than Experiment 2 and 3), (2) some technical issues prevented us from executing further experiments (e.g., overall decay of signals after weeks of experiments or inadvertent removal of headplate etc.), or (3) we deliberately distributed animals across different experiments to ensure enough sample size for each experiment.

A total of 16 animals (Mouse 1-16) were used for the main experiments (GCaMP fluorometry in VS, Experiment 1-3 in the virtual linear track and the moving bar tasks). Four of the 16 animals (Mouse 13-16) had optical fibers implanted both in VS and VTA.

A total of 6 mice (Mouse 13-15, 17-19) were used for GCaMP fluorometry in VTA. Three of the 6 mice had optical fiber also in the VS. Mouse 18 and 19 were further tested with Experiment 5 and 6 (see below).

A total of 6 mice (Mouse 20-25) were used for GFP control experiments.

A total of 6 mice (Mouse 18, 19, 26-29) were used for Experiment 5 and 6 with GCaMP fluorometry in VS. Mouse 29 was further tested with Experiment 4 (see below).

A total of 5 mice (Mouse 29-31, 33, 34) were used for Experiment 4 with GCaMP fluorometry in VS. A total of 4 mice (Mouse 30-33) were used for Experiment 7 GCaMP fluorometry in VS. Note that two mice (Mouse 30 and 31) were tested in both Experiment 4 and 7.

We used an independent set of 8 mice (Mouse 35-42) for closed-loop (operant contingency) experiments.

We used an independent set of 10 mice (Mouse 43-52) for experiments with a dopamine sensor.

We used an independent set of 4 mice (Mouse 53-56) for experiments using rabies virus. One mouse (Mouse 53) did not learn the task before an inadvertent termination of experiments due to COVID19 shutdown. Mouse 53 was, therefore, used only for anatomical investigation. Mouse 54-56 were used for fiber fluorometry experiments as well as anatomical investigation.

A total of 13 mice (Mouse 57-69) were used for electrophysiological recording. Eight of these mice (Mouse 61-66, 68, 69) were additionally examined in the delayed reward task with odor cues. The recording sites of 10 mice (Mouse 57-64, 68, 69) were verified to be in the VTA while those of 3 mice (Mouse 65-67) contained the SNc. We collected the following number of neurons from each animal: 2 neurons from Mouse 57 (medial-lateral coordinate [ML] = 509 μm), 13 neurons from Mouse 58 (ML = 417 μm), 1 neuron from

Mouse 59 (ML = 320 μm), 2 neurons from Mouse 60 (ML = 651 μm), 37 neurons from Mouse 61 (ML = 517 μm), 13 neurons from Mouse 62 (ML = 702 μm), 1 neuron from Mouse 63 (ML = 798 μm), 6 neurons from Mouse 64 (ML = 846 μm), 9 neurons from Mouse 65 (ML = 1048 μm), 5 neurons from Mouse 66 (ML = 936 μm), 6 neurons from Mouse 67 (ML = 1038 μm), 3 neurons from Mouse 68 (ML = 567 μm), and 24 neurons from Mouse 69 (ML = 531 μm).

## SUMMARY OF EXPERIMENTAL ANIMALS

| | Protocol | Animal | Figure |
|---|---|---|---|
| VS, GCaMP | Track 1, standard training | 1-16 | Figures 1K, 1L, and S2E–S2J |
| | No-reward context, Experiment 3, 2, or 1 | 5,6,7,8,9,13,14,15,16 | Figure S4E |
| | Experiment 3 (Speed) | 1-15 | Figures 2I–2L, S3E, and S3F |
| | Experiment 2 (3 teleports) | 1-15 | Figures 2E–2H, S3C, and S3D |
| | Experiment 1 (Teleport and pause) | 5-15 | Figures 2A–2D, S3A, and S3B |
| | Experiment 7 (Reward size manipulation) | 2-11 | Figures 4N–4P |
| | Moving-bar, standard condition | 1,3,5,6,7,8,9,10,13,14,15,16 | Figures S7K and S7L |
| | Moving-bar, Experiment 3 (Speed) | 1,3,5,6,7,8,9,13,14,15,16 | Figures 7I and S7O |
| | Moving-bar, Experiment 2 (3 teleports) | 3,5,6,7,8,9,13,14,15,16 | Figures 7J and S7N |
| | Moving-bar, Experiment 1 (Teleport and pause) | 5,6,7,8,9,13,14,15,16 | Figures 7K and S7M |
| VTA, GCaMP | Track 1, standard training | 13-15, 17-19 | |
| | Experiment 3 (Speed) | 13,14,15,18,19 | Figures 6D and S6G |
| | Experiment 2 (3 teleports) | 13,14,15,17,18,19 | Figures 6C and S6F |
| | Experiment 1 (Teleport and pause) | 13,14,15,17,18,19 | Figures 6B and S6E |
| | Moving-bar, standard condition | 13,14,15,17,18,19 | |
| | Moving-bar, Experiment 3 (Speed) | 13,14,15,17,18,19 | Figure S6J |
| | Moving-bar, Experiment 2 (3 teleports) | 14,15,17,18,19 | Figure S6I |
| | Moving-bar, Experiment 1 (Teleport and pause) | 14,15,17,18,19 | Figure S6H |
| VS, GFP | Track 1, same as GCaMP VS main tasks | 20-25 | Figures 1M, 1N, S4F, and S4G |
| VS, GCaMP | Track 1, training | 18,19,26,27,28,29 | |
| | Experiment 6 (Forward & backward teleport) | 18,19,26,27,28,29 | Figures 4K–4M and S4L |
| | Experiment 5a (Teleport between tracks) | 18,19,26,27,28,29 | Figures 4B–4D and S4I |
| | Experiment 5b (Forward + between-track teleport) | 18,19,26,27,28,29 | Figures 4E–4G and S4J |
| VS,GCaMP | Experiment 4 (Dynamic speed manipulation) | 29,30,31,33,34 | Figures 2M–2P and S3G–S3I |
| | Experiment 7 (Reward size manipulation) | 30,31,32,33 | Figures 4H–4J and S4K |
| VS, GCaMP | Track 1, Close-loop training | 35-42 | |
| | Closed-loop, Track 1, standard | 35-42 | Figure S7C |
| | Closed-loop, Experiment 3 (Speed) | 35-41 | Figures S7E and S7I |
| | Closed-loop, Experiment 1 (Teleport and pause) | 35-41 | Figures S7D and S7H |
| | Closed-loop, Experiment 8a (Scene manipulation) | 36-40, 42 | Figures 7E, S7F, and S7J |
| | Closed-loop, Experiment 8b (Scene manipulation) | 36-40, 42 | Figures 7F, S7G, and S7J |
| VS, DA sensor | Track 1, same as GCaMP VS main tasks | 43-52 | Figures 6E–6H |
| | moving-bar, standard, Experiment 3, 1, 2 | 44,45,47-52 | Figures 7O–7Q |
| VTA, GCaMP | rabies anatomy | 53-56 | |
| (rabies) | Track 1, standard condition training | 54,55,56 | Figure S5J |
| VTA, ephys | Track 1, standard training | 57-69 | |
| | Odor task | 61-66, 68,69 (8/13 mice) | Figures 7A–7D, S7A, and S7B |
| | Track 1, standard training | (65-67: SNc mice) | Figures 5B–5E and S5E |
| | Experiment 1 (Teleport and pause) | | Figures 5K, 5P, and S5O |
| | Experiment 3 (Speed), | | Figures 5L and 5P |
| | Experiment 7 (Reward size manipulation) | | |

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013



### Fiber fluorometry (photometry)

Fluorescent signals from the brain were recorded using a custom-made fiber fluorometry (photometry) system as described in our previous studies (Babayan et al., 2018; Menegas et al., 2017). The blue light (473 nm) from a diode-pumped solid-state laser (DPSSL; 80–500 μW; Opto Engine LLC, UT, USA) was attenuated through a neutral density filter (4.0 optical density, Thorlabs, NJ, USA) and coupled into an optical fiber patchcord (400 μm, Doric Lenses) using a 0.65 NA microscope objective (Olympus). The patchcord connected to the implanted fiber was used to deliver excitation light to the brain and to collect the fluorescence emission signals from the brain. The fluorescent signal from the brain was spectrally separated from the excitation light using a dichroic mirror (T556lpxr, Chroma), passed through a bandpass filter (ET500/50, Chroma), focused onto a photodetector (FDS100, Thorlabs) and amplified using a current preamplifier (SR570, Stanford Research Systems). Acquisition from the red fluorophore (tdTomato) was simultaneously acquired (bandpass filter ET605/70 nm, Chroma) but was not used for further analyses. The voltage signal from the preamplifier was digitized through a data acquisition board (PCI-e6321, National Instruments) at 1kHz and stored in a computer using a custom software written in LabVIEW (National Instruments).

Motion artifacts were examined using mice expressing GFP in dopamine neurons. We injected AAV5-FLEX-GFP in the VTA and SNc of DAT-Cre mice and collected behavior and fluorometry signals in the same way as the GCaMP animals. Raw voltage traces were qualitatively different from GCaMP, with no noticeable fluctuation with the same amplifier configuration. Although animals learned the tasks and acquired behaviors similar to the GCaMP animals, we observed no ramping. We observed neither phasic responses to teleports, nor any significant difference between test conditions.

### Electrophysiology

We based recording techniques on previous studies (Cohen et al., 2012; Kvitsiani et al., 2013; Lima et al., 2009). We recorded extracellularly from the VTA using a custom-built, screw-driven containing eight tetrodes glued to a 200-μm optical fiber (ThorLabs). Tetrodes (Sandvik, Palm Coast, Florida) were glued to the fiber and clipped so that their tips extended 200–500 μm from the end of the fiber. We recorded neural signals with an Open Ephys recording system with Intan headstage (RHD2132, Intan technologies). Broadband signals from each wire were recorded continuously at 30 kHz. To extract spike timing, signals were band-pass-filtered between 300 and 6,000 Hz and sorted offline using MClust-4.3 (A.D. Redish). To be included in the dataset, a neuron had to be well isolated (a measure of unit isolation quality, L-ratio < 0.05) (Schmitzer-Torbert and Redish, 2004). We also histologically verified recording sites by creating electrolytic lesions using 10–15 s of 30 μA direct current.

To unambiguously identify dopamine neurons, we used ChR2 to observe laser-triggered spikes (Cohen et al., 2012; Lima et al., 2009). The optical fiber was coupled with a diode-pumped solid-state laser with analog amplitude modulation (Laserglow Technologies). At the beginning and end of each recording session, we delivered trains of ten 473 nm light pulses, each 5 ms long, at 1, 5, 10, 20 and 50 Hz, with an intensity of 5–20 mW/mm$^2$ at the tip of the fiber. To be included in our dataset, neurons had to fulfill three criteria. (i) The neurons' spike timing must be significantly modulated by light pulses. We tested this by using the stimulus associated spike latency test (SALT) (Kvitsiani et al., 2013). We used a significance value of $p < 0.05$, and a time window of 10 ms after laser onset. (ii) Laser-evoked spikes must be near-identical to spontaneous spikes. This ensured that light-evoked spikes reflected actual spikes instead of photochemical artifacts. (iii) Neurons must have a short latency to spike following laser pulses, and little jitter in spike latency. Post hoc histological analysis showed that the lateral-most recordings penetrated through the SNc. We therefore excluded these recordings from the main analysis. Most of our analyses were based on all of the remaining dopamine neurons recorded in VTA (ML position < 900 μm, 102 neurons, except for Figure 5D).

### Histology

Mice were perfused with phosphate buffered saline (PBS) followed by 4% paraformaldehyde in PBS. The brains were cut in 100-μm coronal sections using a vibratome (Leica). Brain sections were loaded on glass slides and stained with 4′,6-diamidino-2-phenylindole (DAPI, Vectashield). The locations of fiber and tetrode tips were determined using the standard mouse brain atlas (Franklin and Paxinos, 2008).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Theoretical backgrounds

#### *Definitions*

In reinforcement learning theories (Sutton, 1988), the value of a given state is defined as the sum of all future rewards, where rewards are discounted by a constant rate ($\gamma$, discounting factor, $0 < \gamma \leq 1$) per unit time:

$$V(S_t) = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \gamma^3 r_{t+3} + \cdots, \quad (1)$$

where $r_t$ is the reward at time $t$, $S_t$ is the state at time $t$, and $V(S_t)$ is the value of the state $S_t$. Under the assumption that state transitions and rewards follow a Markov process, Equation (1) can be rewritten as:

$$V(S_t) = r_t + \gamma \cdot V(S_{t+1}), \quad (2)$$

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013



which is known as the Bellman equation (Bellman, 1954). The agent approximates the true value $V(S_t)$ with a learned estimate $\widehat{V}(S_t)$, so that if true value is perfectly learned, i.e., $\widehat{V}(S_t) = V(S_t)$, then

$$\widehat{V}(S_t) = r_t + \gamma \cdot \widehat{V}(S_{t+1}) \tag{3}$$

However, before the agent has learned the true value, the left-hand and right-hand sides of Equation (3) will not be equal on average. The difference between these two terms represents the error in value prediction, and as such defines the temporal difference reward prediction error (TD RPE, or $\delta$):

$$\delta_t = r_t + \gamma \cdot \widehat{V}(S_{t+1}) - \widehat{V}(S_t) \tag{4}$$

According to this definition, the TD RPE contains a difference between the estimated values of states that are evaluated at consecutive time points, $\gamma \cdot \widehat{V}(S_{t+1}) - \widehat{V}(S_t)$. When $\gamma = 1$, this term is exactly the temporal derivative of the estimated value function. When $\gamma$ is not 1 but close to it, this term is *approximately* the derivative of estimated value. Thus, the TD RPE is approximately the derivative of estimated value, plus rewards received $(r_t)$(Gershman, 2014). As a result of this property, unexpected increases and decreases in value result in positive and negative transient ("phasic") changes in the TD RPE, respectively.

TD RPEs, as defined by Equation (4), account for three features of dopamine responses in simple classical conditioning paradigms (Schultz et al., 1997):

1. Dopamine neurons are excited by a cue that predicts future reward. In TD models, this occurs because the reward-predicting cue indicates that value at the time of cue presentation is greater than originally expected (i.e., the animal now expects that a reward is coming).
2. When a predicted reward is omitted, dopamine neurons transiently reduce their firing below baseline. In TD models, this occurs because value at the time of omitted reward is now less than originally expected.
3. Unpredicted rewards excite dopamine neurons. However, when a cue predicts delivery of reward, dopamine neurons' response to the predicted reward is greatly reduced. In TD models, this occurs because excitation due to received reward is canceled out by the negative response in 2.

### Conditions for ramping

Consider trials in which a single reward is presented at time $T$. Then before reward is received $(t < T)$, the TD RPE is simply

$$\delta_t = \gamma \cdot \widehat{V}(S_{t+1}) - \widehat{V}(S_t) \tag{5}$$

How must value vary with time in order to produce ramping TD RPEs? We can examine this question by writing the necessary and sufficient condition for a monotonic increase in TD RPEs, i.e.,

$$\delta_{t+1} > \delta_t \tag{6}$$

for all $t < T$. Expanding Inequality (6),

$$\gamma \cdot \widehat{V}(S_{t+2}) - \widehat{V}(S_{t+1}) > \gamma \cdot \widehat{V}(S_{t+1}) - \widehat{V}(S_t), \tag{7}$$

which can be rewritten as

$$\left(1 - \frac{1}{\gamma}\right)\left(\widehat{V}(S_{t+1}) - \widehat{V}(S_t)\right) + \left(\widehat{V}(S_{t+2}) - 2 \cdot \widehat{V}(S_{t+1}) + \widehat{V}(S_t)\right) > 0 \tag{8}$$

The first term on the left-hand side in Inequality (8) corresponds to the temporal derivative of estimated value, scaled by a negative constant, whereas the second term corresponds to the second derivative of estimated value (Mikhael et al., 2019).

Inequality (8) represents the condition for ramping TD RPEs. Note that when $\gamma$ is close to one, the first term is close to zero. Hence, the condition is approximately that the second derivative must be greater than zero, a property that all convex functions satisfy. The exact condition is more restrictive, however: Because value increases with time, the first term is negative. Hence, for the condition to be satisfied, the second term must be positive enough to outweigh the negative first term. Roughly speaking, value functions must be "convex enough" to satisfy the ramping condition.

For an illustration, in Figure S1A we show TD RPEs corresponding to a number of different value functions. It is straightforward to show that value functions in the top panel that are drawn in green satisfy the ramping condition over the evaluated domain, whereas those drawn in red do not. As shown in the bottom panel, only value functions satisfying the ramping condition (Inequality (8)) produce ramping TD RPEs.

### Simulation of speed manipulations

We take states to correspond to locations, and let us assume that the value function is quadratic with states (i.e., $\widehat{V}(S) \sim S^2$). This value function obeys the ramping condition (Figure S1A, green curves). We show in Figure S1B how progressing through states with

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013



variable speed results in value functions of varying convexity (and first derivative) when plotted against time. Applying Equation (4) to these value functions results in TD RPEs that ramp and whose magnitudes increase with speed.

## Quantitative analyses

Error bars and shadings indicate mean ± s.e.m. unless otherwise noted. A significance level of 0.05 was used to determine significant difference. Most of the statistical analyses were performed using a non-parametric test, Wilcoxon signed-rank test unless otherwise noted.

### *Fluorometry (photometry)*

Power line noise in the raw voltage signals was removed by notch filter (MATLAB, Natick, MA). A baseline of the voltage signal was defined by the lowest 10% of signals using a 2-min window. The baseline was subtracted from the raw signal, and the results were z-scored by a session-wide mean and standard deviation (Figures S2A–S2C). For plotting average dopamine responses in Experiment 7 using fiber fluorometry, the fluorescent level from 1-0 s before trial onset was subtracted. Both GCaMP signals and dopamine sensor signals were processed in the same way.

### *Licking and locomotion*

Lick timing was defined as deflection points (peaks) of the output signals above a threshold. To plot the time course of licks, instantaneous lick rate was computed by a moving average using a 200-ms window.

We used the following time window to quantify the lick rate. Impulsive lick: from visual scene movement onset to 2 s before reward; anticipatory lick: [-1 s 0 s] relative to reward; post-reward lick: [0 s 2 s] relative to reward. The same temporal windows are used for quantifying speed. Net anticipatory licking and anticipatory slowdown are defined as below:

$$\text{Net anticipatory licking} = \text{anticipatory lick rate} - \text{impulsive lick rate}$$

$$\text{Anticipatory slowdown} = \text{impulsive speed} - \text{anticipatory speed}$$

### *Dopamine ramping (Ramping R)*

We used the trials in the standard condition to quantify the ramping of fluorescent signals. We computed the Pearson correlation coefficient between time points and z-scored fluorescent signals using either individual trials or averaged response. For the linear track tasks, we used the time window of [-3.5 s −1 s] relative to reward onset for fiber fluorometry data and [-5 s −1 s] for spiking data. We used a smaller window for fluorometry as the slowly delaying response to stimulus or movement onsets often biases the slope of the ramping. For the moving-bar tasks, we used the time window of [-4.5 s −1 s] relative to reward onset.

### *Session-averaged time course*

Lick, locomotion speed, and z-scored dopamine responses for individual trials were aligned by external events (e.g., trial start or teleport onset), and then smoothed using a moving average method. We used a 200 ms time window for licking and spike signals. We did not smooth locomotion speed and fluorometry signals. The results were then averaged across trials for each experimental condition to generate a session-averaged time course.

### *Population-averaged time course*

For fluorometry recording experiments, we typically ran two sessions for each experimental protocol. Since we also did not observe a significant difference in Ramping *R* or Test *R* (see below for definition) between the first and second sessions in the calcium data, we used the second session to plot population-averaged time courses and to perform statistical analyses. We computed the mean of session-averaged time courses from the second session dataset (as the average of all session averages) along with the standard error (the total number of sessions being the sample size) for each experimental condition. Population-average time courses are used to summarize behavior and dopamine responses. We used both sessions for the dopamine sensor data in plotting population PSTHs.

### *Summarizing responses using normalized peak*

Responses to teleport and pause were quantified by peaks of session-averaged time courses using time windows [0.6 s – 2.1 s] and [2 s – 5 s] relative to the teleport and pause onset, respectively. Peaks in the test conditions were normalized by peaks in the standard condition using time window from trial start to reward onset (Figures 2D, left, and 2H, left).

### *Residual responses based on the state value model*

In the teleport conditions, forward scene movement was maintained before and after teleports. As a result, phasic responses to teleports can be contaminated by ramping. For more accurate model comparisons, we generated model responses based on state value predictions (Figures S4A–S4D). For example, if dopamine signals represent RPEs, the deviation from the prediction will be systematically modulated by the conditions: a large positive deviation for long-distance teleport, a small positive deviation for short-distance teleport, and a negative deviation for pause trials. The process begins with a session-averaged time course in the standard condition. Model responses were shifted in time by the amount shortened by teleport (2.2 s for long teleport, 0.25 s for short teleport) or lengthened by pause (5 s). The results were further delayed by 0.3 s to account for neural latency in GCaMP signal. The residual responses were then defined as the subtraction of model response from empirical responses. Deviation from value model was defined as average residual response using time windows [0.6 s – 2.1 s] and [2 s – 5 s] relative to the teleport and pause onset, respectively

Please cite this article in press as: Kim et al., A Unified Framework for Dopamine Signals across Timescales, Cell (2020), https://doi.org/10.1016/j.cell.2020.11.013



(Figures 2D, right, and 2H, right). The baseline activity for each trial, defined by average response using time window [–1 s - 0 s] from trial start, was subtracted from the average residual responses.

### Summarizing the change of responses (ΔF/F difference)

To quantify the change of calcium responses at the time of teleport in Experiment 5c (Figure 4J) and Experiment 6 (Figure 4M), we first computed mean responses using a pre-teleport window (−0.2 s and 0 s relative to teleport) and a post-teleport window (0.6 s and 2.1 s relative to teleport) for each session. We then subtracted the responses using the first window from the responses using the second window.

### Summarizing responses for single unit data

For individual trials, responses to reward, teleport, pause were quantified by averaging responses using time windows [0.05 s - 0.45 s], [0.1 s - 0.5 s], [0.1 s - 0.5 s] relative to the events of interest, respectively. Baseline responses were defined by average responses using time window [–1 s - 0 s] relative to trial start. The baseline responses were subtracted from the responses of interests to obtain a net modulation by the events.

### Estimation of GCaMP signals from spiking data

We reasoned that the key difference may lie in the different kinetics between spike and calcium signals. That is, a single spike produces a seconds-timescale increase in calcium due to the slow dynamics of calcium and the kinetics of calcium indicators (GCaMP6m). To test our hypothesis, we generated predicted calcium signals based on the spiking activity obtained in our recording experiment and an impulse response of GCaMP6m to a single spike (Figure S5P). We estimated GCaMP responses of a single neuron based on a relationship between a single spike and GCaMP response (Chen et al., 2013b). We convolved the spike train from the entire session with the GCaMP kernel (Figure S5P) to obtain estimated GCaMP signals from a single neuron activity (Figure 5F). GCaMP signals measured from fiber fluorometry can be approximated by calcium signals pooled across neurons. To estimate the fluorometry response of a single trial, we randomly selected single trials of convolved responses in the same experimental condition aligned by the event of interest and summed up the convolved responses across neurons. We repeated this for multiple trials (n = 100 for Figure 5F; n = 200 for Figures 5K and 5L) to generate predicted GCaMP responses.

The results were then transformed into z-scores. For Figure 5F, the z-scored responses were baseline-subtracted for each subgroup of neurons. It should be noted that the transformation into z-scores effectively discards the information about baseline firing. This is a part of the reasons why ramping signals can be exaggerated relative to the baseline. We note that the quantification of fluorometry signals are often based on relative changes from baseline, with the baseline defined as the level of fluorescence in a certain time window. This method also discards the information about the level of baseline activity, and may exaggerate ramping signals such as those observed in the present study.

### Summarizing responses in the test condition (Test R)

The systematic deviation from value model across conditions was quantified using Spearman correlation for each session. For Experiment 1 (teleport and pause), numbers were assigned to each condition to quantify the trial-by-trial correlation between test conditions and residual responses based on the value model (long teleport = 1; short teleport = 2; pause = 3). Significant negative correlations indicate that residual responses are big in the long teleport, medium in the short teleport, and the smallest in the pause condition. Alpha = 0.05 was used to determine significanace in Test Rs. For Experiment 2 (three teleports), we computed Spearman correlations between position and the residual responses. Positive correlations indicate that responses to teleports increase with proximity to reward location. For Experiment 3, we computed Spearman correlation between speed condition and baseline-corrected (using time window [–1 s – 0 s] from trial start) average response from trial start to reward onset. Numbers were assigned such that positive correlations indicate that responses increase with scene speed (slow speed = 1; standard speed = 2; fast speed = 3). In addition to testing the significance of trial-by-trial correlation in an individual session, we further tested whether the median of Test *R*s across animals is significantly different from zero using Wilcoxon signed-rank test.

### Statistical analysis

We performed statistical analyses at both single-session and population levels. For individual-session analysis, average responses from individual trials were quantified using a temporal window locked to an external event (e.g., reward onset). Non-parametric tests (e.g., Wilcoxon signed-rank test) were used to test whether responses are significantly greater or smaller than reference (e.g., zero). A dataset with a significant difference was marked by a filled circle. For population-level analysis, the mean responses of individual sessions were used for comparisons between conditions or comparisons against a reference value (e.g., zero or one) using Wilcoxon signed-rank test. A significant difference between conditions was marked by a filled circle on the horizontal lines that connect the two conditions. A significant difference from baseline was marked using a p value or a star mark on top of each condition. We used two-tailed tests for all statistical tests.

### Across-session Ramping R analysis

We examined whether and how much task-related behaviors can account for the variability in Ramping *R* across sessions. For each session, we quantified average Ramping *R*, net anticipatory lick rate, and running speed. We examined whether lick rate or running speed can account for the variability in Ramping *R* using analysis of covariance (ANCOVA). In the analysis, Ramping *R* was dependent variable, either lick rate or running speed was covariates, and animal identity was categorical independent variable.

### Trial-to-trial regression analysis for Experiment 3

We examined how trial-to-trial variability in the mean dopamine responses during approach can be accounted for by visual scene speed, locomotion speed, and a global trend by multiple linear regression (Figure 2L, right).

$$Response = \beta_0 + \beta_{vis}V_{vis} + \beta_{loc}V_{loc} + \beta_{trial}Trial_{number} + \varepsilon$$

## Model fitting

### *Preprocessing of data and kernel for model fitting*

For fluorometry data, z-scored dopamine signals and other behavior signals (stored at 1kHz) were downsampled to 100Hz and then averaged for each condition aligned by trial start. We first fitted individual sessions independently (Figure 3C). For combined dataset (Figure 3D, 'all conditions'; Figures 3E–3H), we concatenated data from different experiments. This allowed us to fit the data from the same animal using a single set of parameters. Data from Experiment 4 were not combined as the data were collected using different animals (Figure S4H). Since the number of animals is not large, data were included in the combined dataset (Experiments 1-3) as long as the animal has at least one manipulated experiment. We obtained the kernel for GCaMP ($F$) by averaging GCaMP responses to a delivery of unexpected reward during the last day of training (Figure S5P). The results were similar when we used a single-spike response filter (Chen et al., 2013b) that we used for predicted calcium response (Figure 5E).

For spiking data, responses were fitted for each neuron. We used neurons that have both Experiment 1 and Experiment 3 (n = 78 neurons). Due to the low firing rates of dopamine neurons (~5 spikes/s), time courses of firing rate using a 100ms window were often noisy and caused low $R^2$. We further smoothed it using a 250 ms window. The results were concatenated across experiments, and then z-scored. Z-scoring enabled us to use the same boundary condition for almost all models parameters as what we used for fitting fluorometry data. For the kernel filter ($F$), we used average response to unexpected reward (Figure S5P).

### *Model fitting procedure*

For quantitative comparisons between the state value and RPE models (Figure 3), we examined which model provides a better fit to the mean calcium signals from each animal. For this analysis, we focused on the pre-reward period where ramping dopamine signals were observed. We first defined the shape of the value function across space. We then predicted the time course of value or TD RPE signal in each experimental condition. We then converted the predicted value or TD RPE into either calcium signals or firing rates, and these predictions were compared with the mean responses in the data.

The TD RPE at time $t$ $(\delta_t)$ is defined by:

$$\delta_t = \gamma \cdot \widehat{V}(S_{t+1}) - \widehat{V}(S_t)$$

where $S_t$ is the state at time $t$, $\widehat{V}(S_t)$ is the estimated value of the state $S_t$, and $\gamma$ is the discounting factor $(0 < \gamma \leq 1)$. Note that this formulation excludes reward delivery times. To fit responses in the linear track tasks, we used the position along the linear track as $S_t$; that is, the state value is defined as the value of the position. To fit responses in the moving-bar tasks, we used the vertical position of the bar as $S_t$.

The state value is expected to increase as the animal gets closer to the goal location. For the TD signal to be positive and ramp up, the value function must take a convex shape along the relevant dimension along which the animal traverses (See STAR Methods, Theoretical backgrounds). Because the exact shape of the state value is unknown, we examined several shapes of value functions. Our first model defines the shape of the value function as an exponentially decaying function across space with the discounting factor $\tau$.

$$\widehat{V}(S_t) = \beta_1 \cdot \tau^{(S_T - S_t)}$$

where $\beta_1$ is a coefficient representing the value at the goal location while $S_T$ is the position of the goal (target), and therefore, $S_T - S_t$ corresponds to the distance from the current position to the goal. Note that the discounting factor $\tau$ need not be the same as the discounting factor $\gamma$ for the definition of TD errors.

We also fitted the measured data using a time-to-reward model (Δt-to-reward). In this model, the value function is computed based on the estimated time to obtain the reward. The estimation is done by dividing the distance to reward by the instantaneous speed (smoothed by a 100 ms window).

$$\widehat{V}(S_t) = \beta_1 \cdot \gamma^{(\Delta t)}$$

where $\Delta t = (S_T - S_t)/v_t$. Note that this model produces some features that other value models do not. For example, when the animal is paused in Experiment 1, Δt becomes infinite and the value at the moment becomes zero. When the animal is teleported to the closer position, the instantaneous speed suddenly increases, which results in a transient increase of value. Although the model predictions in Experiment 1 somehow mimic those in the TD models, the predicted response in Experiment 3 does not increase with speed, resembling other value models.

Latent varables (value and TD RPE) in our models are computed based on the instantaneous position sampled at 100 Hz. Due to the perceptual delay as well as the intrinsic biological mechanisms underlying the signals (e.g., kinetics of action potential or intracellular calcium), The dynamics of measured neural signals are slower than the dynamics of latent variables in our models. To account for the difference in the temporal dynamics, the predicted value or TD RPE was convolved with a kernel filter ($F$) to reflect the kinetics of

measured signals. The filter was estimated from the empirical impulse function in response to uncued reward as described above. For the TD RPE model, the calcium signals $y_t$ were predicted by convolving $\delta_t$ with the kernel filter and an offset term.

$$y_t = (\beta_0 + \delta_t) * F + \epsilon_t$$

where $\varepsilon_t$ denotes Gaussian-distributed random error. For the value model, calcium signals were predicted by convolving $\widehat{V}(S_t)$ with the kernel filter.

$$y_t = \left(\beta_0 + \widehat{V}(S_t)\right) * F + \epsilon_t$$

We also fitted the measured data using a RPE-value mixture model. This model is a weighted linear sum of the RPE and value terms. The coefficient for the RPE model reflects the relative contribution of RPE when both models were used to explain the data.

$$y_t = \left(\beta_o + \alpha Z(\delta_t) + (1-\alpha) Z\left(\widehat{V}(S_t)\right)\right) * F + \epsilon_t$$

where Z(x) denotes z-scoring. We noticed that occasionally the best fit results consist of $\delta_t$ and $\widehat{V}(S_t)$ that were not interpretable (e.g., positive $\widehat{V}(S_t)$ and negative $\delta_t$ whose mangitudes were much larger than the measured signals). To avoid this, we applied a magnitude constraint on $\widehat{V}(S_t)$ by adjusting the cost function. If the range of convolved $\widehat{V}(S_t)$ in the standard condition $(L_{\widehat{V}})$ was larger than twice of the range in the measured signal $(L)$, the sum of squared residuals (SSR) was increased by a multiplier as shown below. This constraint made each component reasonable without a noticeable effect on the goodness-of-fits (r square decreased on average by 0.06% for fluorometry data, 0.03% for spike data).

$$\text{Adjusted SSR} = \text{SSR} * \frac{L_{\widehat{V}}}{L} * 100, \quad \text{if} \quad L_{\widehat{V}} > L$$

Finally, we fitted the measured data using a model in which the derivative nature of the model itself can vary from being a zeroth order derivative (equivalent to the value model) to a second order derivative (a model with a first order derivative is approximately the RPE model) in a continuous manner. Fractional derivatives can be defined to extend differential calculus with integer numbers to non-integer numbers. We used the MATLAB function *fgl_deriv* that computes the fractional derivatives using Grünwald–Letnikov formulation (Podlubny, 1998).

$$y_t = \left(\beta_o + \mathit{fgl\_deriv}(\widehat{V}(S_t), \alpha)\right) * F + \varepsilon_t$$

A model fit was performed by minimizing the sum of squared residuals (SSR) in the time window from the scene movement onset to 0.5 s before reward delivery. To find parameters that minimize SSR numerically, we used a non-linear function solver with constraints (fmincon, MATLAB). Sets of the example starting point (p0), lower bound (LB) and upper bound (UB) are as follows:

- TD model, exponential $(\gamma, \tau, \beta_1, \beta_0)$: $\widehat{V}(S_t) = \beta_1 \cdot \tau^{(S_T - S_t)}$, $\delta_t = \gamma \cdot \widehat{V}(S_{t+1}) - \widehat{V}(S_t)$

p0 = [0.96 0.96 25 0]; LB = [0.8 0.8 1 –2]; UB = [1 1 150 2];

- Value model, exponential $(\tau, \beta_1, \beta_0)$: $\widehat{V}(S_t) = \beta_1 \cdot \tau^{(S_T - S_t)}$

p0 = [0.96 0.6 0]; LB = [0.8 0 –2]; UB = [1 10 2];

- Value model, time-to-reward $(\tau, \beta_1, \beta_0)$

p0 = [0.5 0.6 0]; LB = [0.1 0 –2]; UB = [1 10 2];

- TD-value mixture model, exponential $(\gamma, \tau, \beta_1, \beta_0, \alpha)$: $\widehat{V}(S_t) = \beta_1 \cdot \tau^{(S_T - S_t)}$

p0 = [0.96 0.96 25 0 0.5]; LB = [0.8 0.8 0.005 –2 0]; UB = [1 1 500 2 1];

- fractional derivative model, exponential $(\tau, \beta_1, \beta_0, \alpha)$

p0 = [0.95 0.6 0 0.5]; LB = [0.8 0 –2 0]; UB = [1 200 2 2];

Starting points of parameter search were either the example starting point (p0) or randomly drawn from a uniform distribution ranging between the lower and upper bounds. The solver was repeated 250 times (50 times for the fractional derivative), and the parameters with the minimum SSR were chosen.

Empirical value functions may deviate from a simple mathematical form (e.g., exponential). To model more general shapes of the value function, we also used a fifth order polynomial regression.

$$\widehat{V}(S_t) = \sum_{k=0}^{5} \beta_k S_t^k$$

$$\delta_t = \gamma \cdot \widehat{V}(S_{t+1}) - \widehat{V}(S_t) = \sum_{k=0}^{5} \beta_k \left(\gamma S_{t+1}^k - S_t^k\right)$$

$$y_t = \widehat{V}(S_t) * F + \epsilon = \sum_{k=0}^{5} \beta_k \left(S_t^k * F\right) + \epsilon$$

$$y_t = \delta_t * F + \epsilon = \sum_{k=0}^{5} \beta_k \left(\left(\gamma S_{t+1}^k - S_t^k\right) * F\right) + \epsilon$$

The coefficients of the value model can be found deterministically. For the TD model, we performed regression using a range of $\gamma$ values from 0.97 to 1 in steps of 0.0005 and used the $\gamma$ value with the maximum $R^2$.

Since polynomial regression allows the value function to decrease as animals approach, we further used a constrained linear least square method ('lsqlin' function in MATLAB) to make the value function monotonically increasing. V(s + 1) – V(s) > 0, s = 0, 2, 3, … 94. For the TD model, SSR were computed over a set of $\gamma$ values, and the best parameters were chosen as we did for the regression model.

To compare goodness-of-fits across models, Akaike information criterion (AIC) (Akaike, 1973) was used.

$$AIC = N \times \ln\left(\frac{SSR}{N}\right) + 2K$$

where $N$ is the number of data points, SSR is the squared sum of residuals, $K$ is the number of free parameters. AIC was computed using the time window from the scene movement onset to 0.5 s before reward delivery. When computing AIC for polynomial fits, the number of parameters ($K$) was increased by one to compensate for the exhaustive search for $\gamma$. For comparisons, we computed differential AIC between a target model and a reference model ($\Delta AIC = AIC_{target} - AIC_{ref}$). Positive value indicates that the reference model explains the data better than the target model considering the number of parameters.

Permutation tests were used to test significance of R square and AIC difference. The data were randomly shuffled using 1 s bins and the same fitting procedures were performed to find the best fit. We repeated it 1000 times to obtain a distribution of goodness-of-fits. To test significance of R square, P value was defined as the faction of R squares from shuffling that is greater than the R square from actual (unshuffled) data. To test significance of AIC difference, a distribution of shuffled AIC difference was obtained by subtracting the AICs of the exponential TD model from the AICs of the model of interest. P value was defined as the fraction of shuffled AIC difference whose values are more extreme than the empirical AIC difference. Alpha = 0.05 was used to determine significance. Fractional derivative model was omitted for significant test as the fractional derivatives computation took orders of magnitude longer than other fittings thus it was not feasible to perform a permutation test.

We reconstructed the shape of value function based on the the optimal fit parameters in the fit using polynomial basis function with monotonically increasing constraint (Figures 3F and 5F). We first generated the time course of position in the standard condition, and the value function was computed based on the position at each time point. The results was normalized by the peak for each animal (flurometry) or neuron (spike). RPE signals were computed based on the reconstructed value function and the optimal gamma. The results were averaged across animals (Figure 3F, right, dark green) or neurons (Figure 5R, right, dark green). GCaMP filter (Figure S5P, green) was convolved with the RPE signals and the results were averaged across animals (Figure 3F, green) or neurons (Figure 5F, green). For smooth visualization of the predicted signals, the long-tail part of the GCaMP filter (Figure S5P, green; 4 s to 5 s) was multiplied by a linear function, y = (5 – x).

### Note on the shape of the fitted value function

First, the RPE model contained the initial transient response that corresponds to a stepwise increase of value at the trial start (Figure 3F, right). These responses resemble dopamine neuron responses to cues that predict reward or a trial start signal (Bromberg-Martin et al., 2010; Schultz et al., 1997). Second, the obtained value function showed that the convexity of the value function decreases closer to the goal (Figure 3F, left), which results in a decrease in TD errors (though, it is still positive) and, the corresponding calcium signals immediately before reward delivery (Figure 3F, right). A similar change in value function as well as a slight decrease in dopamine signals before reward was observed in previous studies, which suggested the importance of state uncertainty in generating these changes (Cohen et al., 2012; Starkweather et al., 2017).

# Supplemental Figures

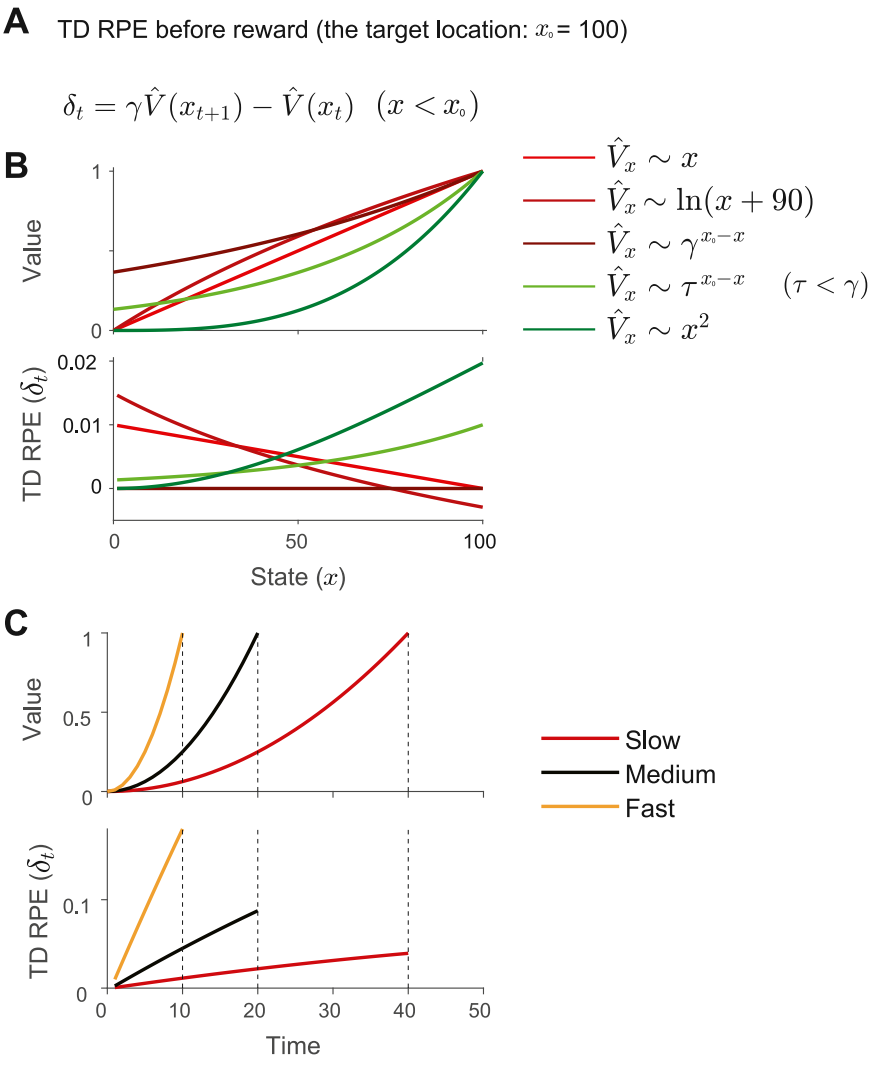



**Figure S1. Relationship between Value Function and TD RPE Signal, Related to Figure 1**

(A) Equation for TD RPE ($\delta$) before the reward location ($x_0 = 100$). $\hat{V}_x$, estimated value of state $x$. $\gamma$, discounting factor ($0 < \gamma \leq 1$). (B) Conditions for which TD RPE can ramp up or down. (top) Value functions. (bottom) TD RPEs. Five different forms of value functions and resulting TD RPEs following the equation in (A) are shown. The value functions drawn in green satisfy the ramping condition over the evaluated domain, whereas those drawn in red do not. Note that when the value is discounted by the discounting factor $\gamma$, TD RPE is zero (dark brown). By contrast, when the value becomes *convex enough* due to a separate spatial discounting factor $\tau$ ($\tau < \gamma$), TD TPE can ramp up (light green). Here we set $\gamma = 0.99$, and $\tau = 0.98$. See STAR Methods, Theoretical backgrounds. (C) Relationship between speed and TD RPEs. When states are traversed faster, the convexity of the value function will be accentuated when plotted against time. Hence, TD RPEs show greater ramping in faster conditions. Dashed black lines denote the end of the trial for the fast, medium, and slow conditions. Here, we set $\gamma = 0.99$, track length to 20, and the speeds to 2, 1, and 0.5, for the fast (yellow), medium (black), and slow conditions (red), respectively. Value takes a quadratic shape, and the peak value is normalized to one, i.e., $\hat{V}(x) = \left(\frac{x}{20}\right)^2$.

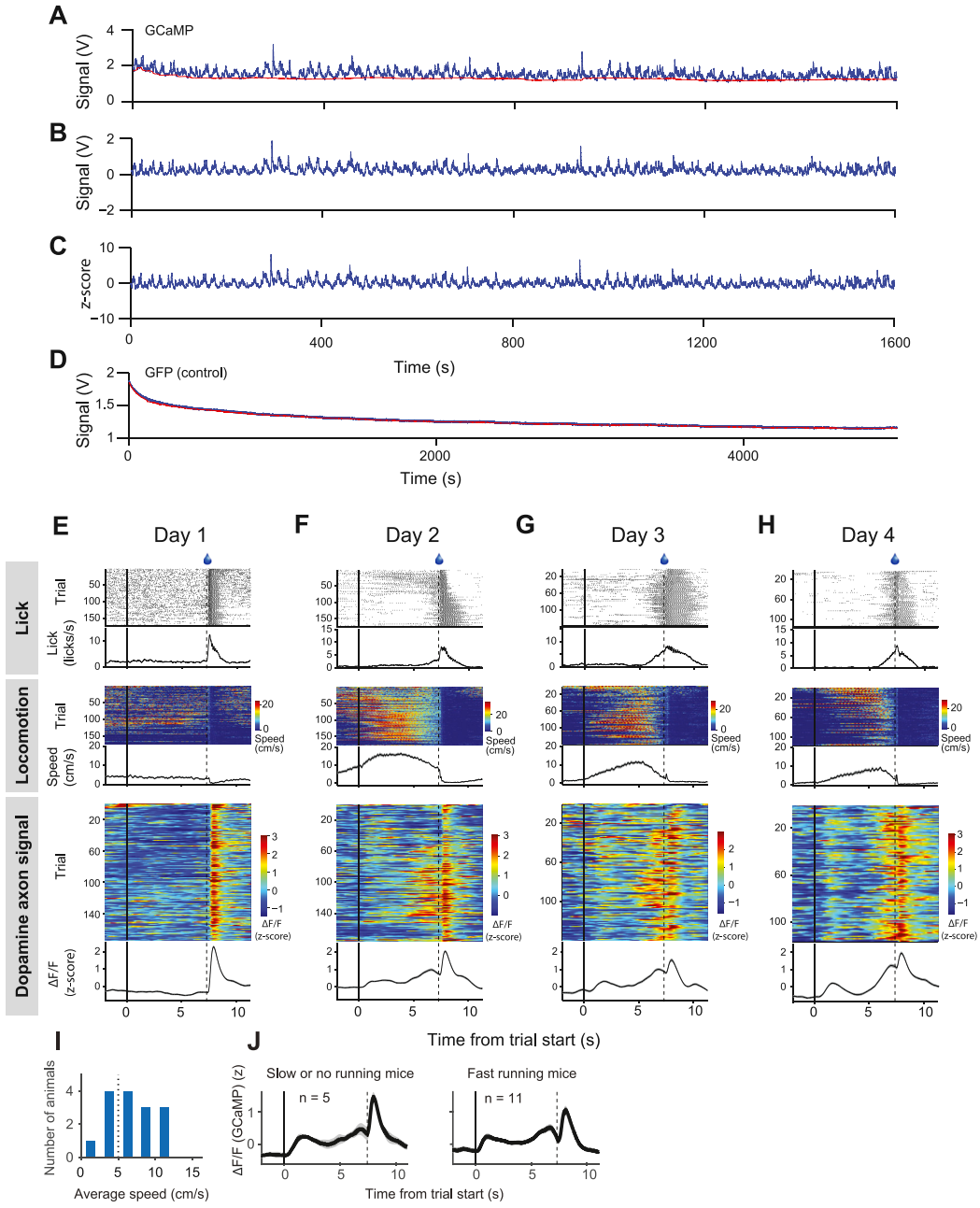



**Figure S2. Fluorometry Signal Processing and Example Training Sessions, Related to Figure 1**

(A) Raw voltage outputs from a current amplifier attached to a photodiode (blue). Slow drift (red) was defined by the lowest 10% of signals using a 2-min moving window. Drift-corrected signal (B) was defined by subtraction of slow drift baseline (A, red) from the raw signals (A, blue). (C) A session-wide mean was subtracted from drift-corrected signals, then the result was divided by the session-wide standard deviation to calculate z-scored fluorescent signals. (D) Raw signals collected from a control animal expressing GFP in dopamine neurons. Signals from the same amplifier gain as (A) are very smooth and lack fluctuations, confirming that motion artifacts were negligible in our head-fixed setup. (E) Data collected on day 1 of training. (top) Raster plot showing lick events aligned by scene movement onset. Lick events are averaged across trials using a temporal window of 0.2 s. (middle) Instantaneous speed is color-coded and averaged across trials. (bottom) Z-scored dopamine axon signals are color-coded and averaged across trials. (F), (G), (H). Data collected on days 2, 3, and 4 (the last day of training for this animal), respectively. Format as (E). (I) Summary of average running speeds during the task on the last day of training sessions (n = 16 mice). (J) Time course of z-scored GCaMP signals from animals that showed slow or no running (left, n = 5 mice; running speed $\leq$ 5cm/s) and those that showed fast running (right, n = 11 mice; running speed > 5cm/s). The median Ramping *R*s were not significantly different between the two groups (p = 0.27, n = 5 and 11, Wilcoxon rank-sum test).

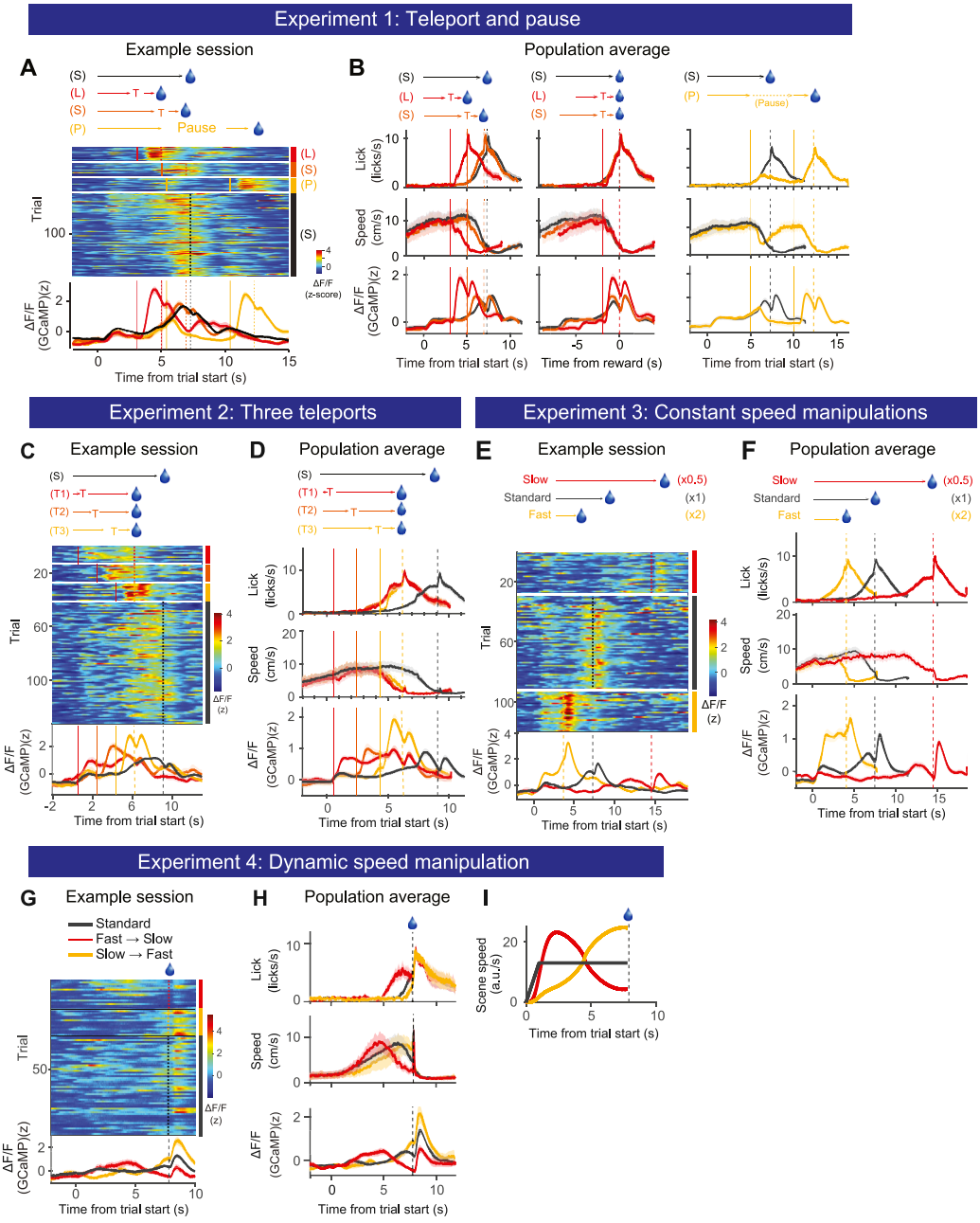



**Figure S3. Example Calcium Recording Sessions and Population Average for Experiments 1, 2, 3, and 4, Related to Figure 2**

(A) An example session for Experiment 1 (teleport and pause experiment). (top) Time courses of events for each condition (black S: standard; red L: long teleport; orange S: short teleport; yellow P: pause) in Experiment 1. (middle) Z-scored dopamine axon signals from an example session. Trials are sorted by conditions; long-distance teleport (L, red), short-distance teleport (S, orange), pause (P, yellow), and the standard condition (S, black). (bottom) Responses are averaged across trials for each condition. (B) Time courses of the average lick (top row), locomotion speed (middle row), and dopamine axon signals (bottom row) across animals (n = 11 mice). Teleport responses are aligned by scene movement onset (left column) or reward onset (middle column). Note that anticipatory lick and slowdown of locomotion overlap when responses are aligned by reward onset, suggesting that animals' appetitive behaviors were based on their position in the virtual space but not based on the elapsed time alone. (right column) Responses in the pause condition aligned by scene movement onset. (C) An example session for Experiment 2 (Three-teleport experiment). Trials are sorted by teleport at short (red), middle (orange), or long (yellow) distance from the start location. Black indicates the standard condition. (D) Average lick, locomotion speed, and dopamine axon signals (n = 11 mice). Teleport onsets (solid line) and water delivery (dashed line) are marked. (E) An example session for Experiment 3 (Speed-manipulation experiment). Trials are sorted by the speed of scene movement (× 0.5: red, × 1: black, × 2: yellow). (F) Population-average of lick, locomotion speed, and dopamine axon signals (n = 15 mice). (G) An example session for Experiment 4. Trials are sorted by the speed profile. (H) Average lick, locomotion speed, and dopamine activity (n = 5 mice). (I) Time courses of speeds for each condition.

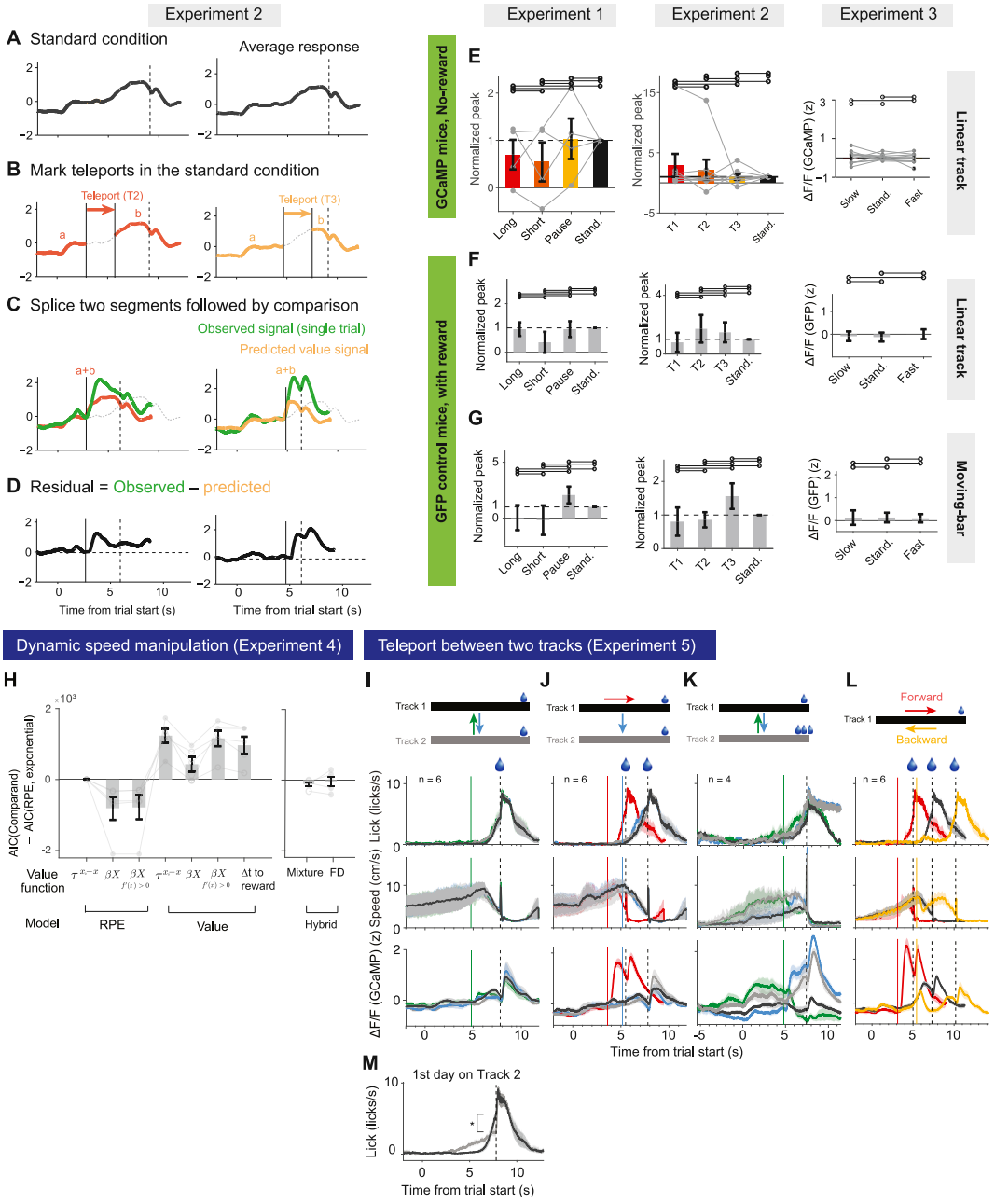



**Figure S4. Quantification of Residual Responses and Data from the Control Conditions, Related to Figures 3 and 4**

(A-D) Step-by-step procedures for computing residual responses of value model using examples from Experiment 2 (T2 and T3 trials). (A) Average response in the standard condition (black) was used to predict value signals in the experimental conditions, assuming that the value is a function of location $(x)$. (B) The start and end locations of a teleport were converted into the time points in the standard condition (black line). Predicted value signals were obtained by splicing (C) the segments before and after the teleport timing (a and b, respectively in B). The predicted value signal was then compared with the observed signal. (D) The residual was obtained as the difference between the observed and predicted value signals. (E) Results from no-reward control sessions. Prior to the beginning of training for the standard task, a subset of animals performed Experiments 1-3 as shown in Figure 2 but without reward at the target location (see STAR Methods). (left) Peaks for each condition ($p = 1.00$, $n = 4$ mice, Kruskal-Wallis test). (middle) Normalized peaks for each condition ($p = 0.33$, $n = 8$ mice, Kruskal-Wallis test) in Experiment 2 (three-teleport experiment, corresponding to Figure 3G). (right) Mean responses at [–1 s 0 s] relative to reward onset ($p = 0.65$, $n = 8$ mice, Kruskal-Wallis test) in Experiment 3 (speed-manipulation experiment). (F, G) Results from GFP-control animals. Motion artifact in the fluorometry signals was assessed by GFP control animals (See Methods). (F) Summary of results in Experiments 1-3 in the linear track tasks ($n = 5$ mice). (G) Summary of results in Experiments 1-3 in the moving-bar tasks ($n = 4$ mice). (H) AIC comparisons between value, RPE (left), and hybrid models (right) in Experiment 4 (dynamic speed). (I) Teleport between tracks experiments. In the standard conditions, animals completed trials in Track 1 (black) or Track 2 (gray). In the between-track teleport conditions, animals were teleported from Track 1 to Track 2 (cyan), or teleported from Track 2 to Track 1 (green) ($n = 6$ mice). (J) Animals were either teleported forward (red) in Track 1

*(legend continued on next page)*

or teleported to Track 2 (cyan) (n = 6 mice). (K) Animals obtained small (2μL) and large (12μL) reward in Track 1 (black) and Track 2 (gray), respectively (n = 4 mice). (L) Average lick, locomotion speed, and dopamine responses in the forward (red) and backward (yellow) teleports experiments (n = 6 mice). (M) Average lick on the first day introducing Track 2 for Experiment 5a (l). Anticipatory licks right before reward are significantly different between the two tracks ($p = 0.031$, n = 6 mice, Wilcoxon signed-rank test), showing that the aniamls could perceptually distinguish the two tracks.

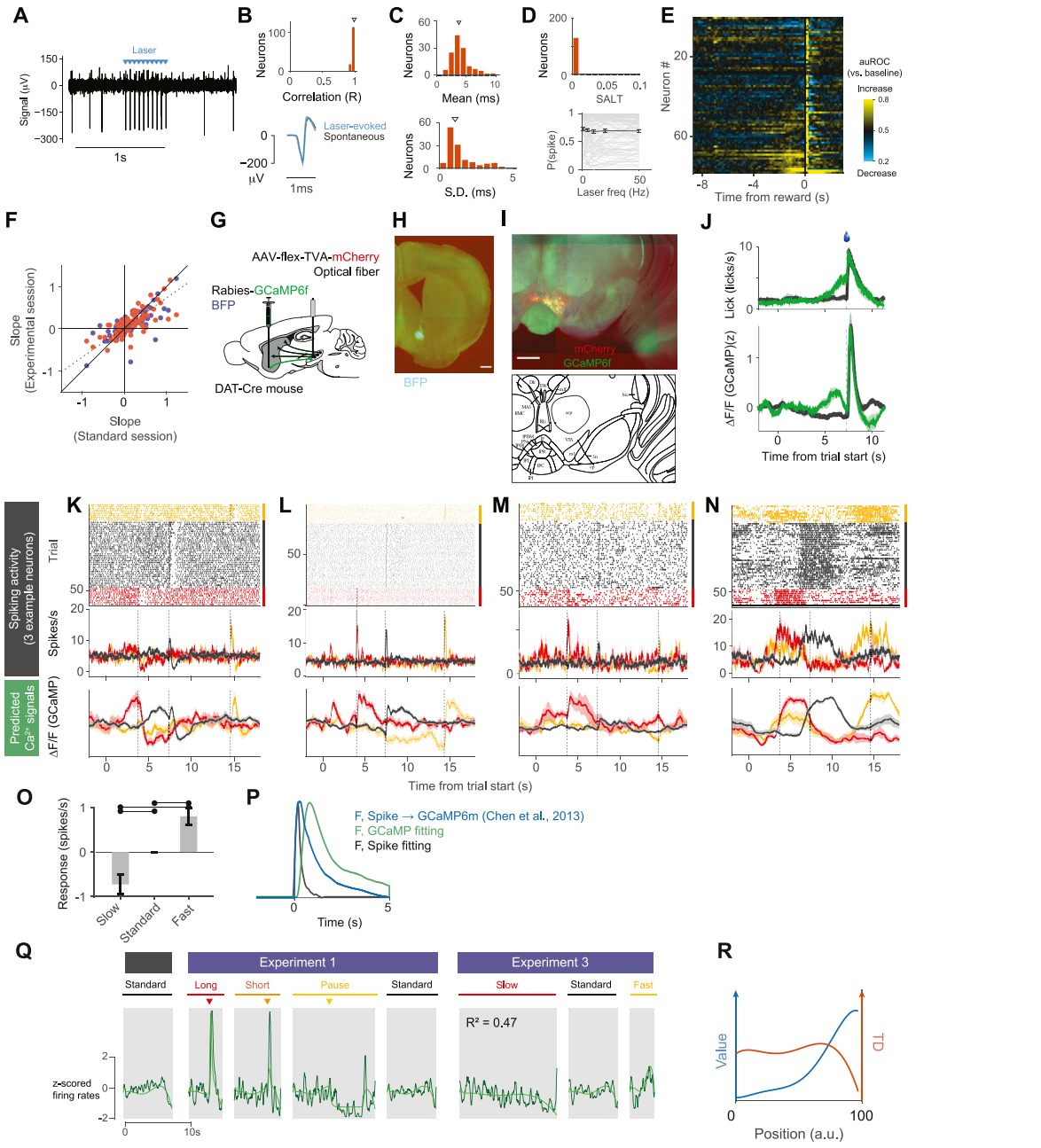
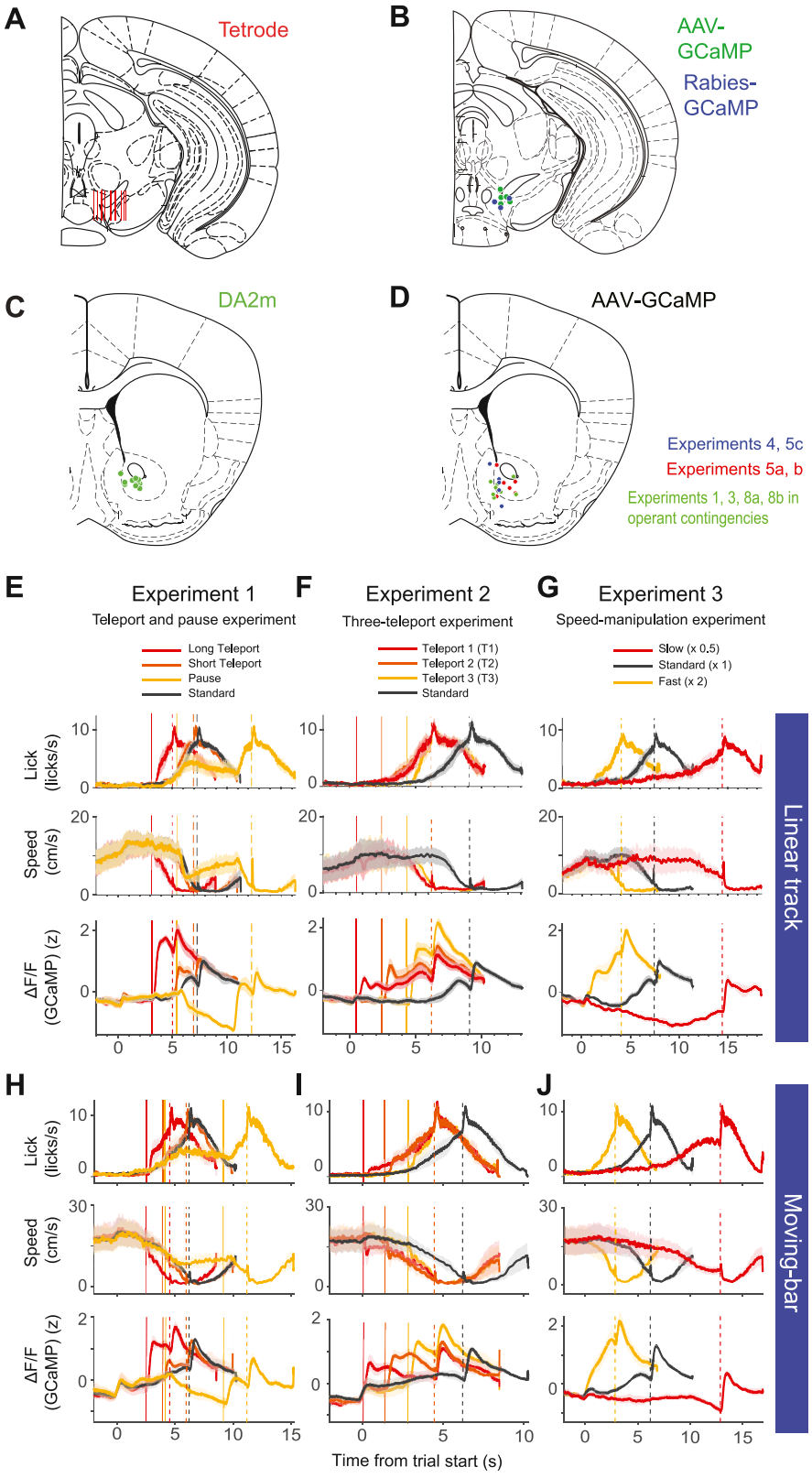



**Figure S5. Spiking of Dopamine Neurons with Heterogeneous Ramping Encode RPEs, Related to Figure 5**

(A) The raw voltage signal collected during optogenetic identification of dopamine neurons. This neuron reliably responded to laser pulses (cyan, 20Hz, 5ms duration). (B-D) Characterizations of optogenetically tagged dopamine neurons (n = 131 neurons, including 20 lateral neurons only included in Figure 5D). (B) (top) Histogram of Pearson correlation coefficients between laser-evoked and spontaneous spike waveforms. (bottom) Comparison of spike waveforms from a single neuron between laser-evoked spikes (cyan) and spontaneous spikes (black). (C) Latency of laser-evoked spikes. (top) Histogram of average latencies across neurons. Triangle denotes the median (3.77 ms). (bottom) Histogram of the standard deviations of latencies across neurons. Triangle denotes the median (1.06 ms). (D) (top). All of the identified neurons were significantly modulated by laser pulses ($p < 0.05$, SALT test) (Kvitsiani et al., 2013). (bottom) Probability of spike occurrence for each laser frequency. (E) Responses (auROC against baseline) sorted by the slope of ramping (n = 78 neurons). (F) The slope of ramping was preserved across different sessions of experiments (blue, $r = 0.77$, $p = 3.5 \times 10^{-17}$, Experiment 1; red, $r = 0.76$, $p = 1.8 \times 10^{-16}$, Experiment 3, Spearman correlation). In each day, a session that contains only the standard condition (standard session) was followed by sessions with Experiment 1 and Experiment 3. Mean of the slopes for each animal were subtracted to parse out the medial-lateral effect (Figure 5D). (G) An illustration of Rabies injection experiment. AAV-flex-TVA-mCherry was injected in VTA and Rabies-GCaMP and BFP was injected in VS, targeting nucleus accumbens core (see STAR Methods). (H, I) Representative images showing focal expression of BFP, which indicates the injection site, around nucleus accumbens (H), and the expression of mCherry (red) and GCaMP6f (green) in the VTA (I). The histological images were created by stitching together multiple images with a higher resolution. (J) Average time courses of licking (top)

*(legend continued on next page)*

and dopamine calcium activity in VTA (bottom) in the early (black) and late (green) training (n = 3 mice). (K-M) Spiking activity of the positive-ramping (K; Figure 5G), negative-ramping (L; Figure 5H), and no-ramping (M; Figure 5I) neurons in Experiment 3. (N) Responses of the neuron in Figure 5J. (O) At the single neuron level, neuron's firing rates increased with the speed of the scene ($p = 5.5 \times 10^{-7}$ and $8.3 \times 10^{-5}$, n = 83 neurons, Wilcoxon signed-rank test for slow speed versus standard and fast speed versus standard, respectively). (P) Spike filter (black) and GCaMP filter (green) used for model fitting analysis. These filteres were defined by average responses to uncued response to reward (see STAR Methods). Impulse response filter used to compute predicted GCaMP from spikes (blue) (Chen et al., 2013b). (Q) An example of data (dark green) and prediction from the fitting (light green) using a polynomial basis with monotonically increasing constraint ($R^2 = 0.47$). (R) The shape of the value function reconstructed using the optimal parameters in (Q) (blue). TD signal (red) was computed based on the reconstructed value function, the optimal temporal discounting factor $\gamma$, and the movement speed in the standard condition.




*(legend on next page)*

**Figure S6. Dopamine Cell Body Activity in the VTA Encodes RPE, Related to Figures 5 and 6**

(A) Tetrode positions for single-unit recordings in the VTA (Figure 5). (B) Fiber tip positions for cell body calcium recordings in the VTA. Green, AAV-GCaMP injection (n = 6 mice). Blue, Rabies-GCaMP injection (n = 3 mice). (C) Fiber tip positions for dopamine concentration recordings in the VS. (D) Fiber tip positions for calcium recordings in the ventrial striatum. Blue, positions for Experiment 4 and 5c (n = 5 mice; Figures 2M–2P and 4H–4J). Red, positions for Experiment 5a, 5b, and 6 (n = 6 mice; Figures 4A–4G). Green, positions for Experiments 1, 3, 8a, and 8b in operant contingencies (n = 8 mice; Figures 7E–7G; Figures S7C–S7G). (E-G) Average lick (top), locomotion speed (middle), and fluorometry (photometry) signals in VTA (bottom) in Experiment 1 (E), n = 6 mice), Experiment 2 (F), n = 6 mice), and Experiment 3 (G), n = 5 mice) in the virtual linear track experiments. Vertical lines represent teleport or pause events. Dotted lines represent reward onsets. Format same as Figures 2C, 2G, and 2K, respectively. (H-J) Results in Experiment 1 (H), n = 5 mice), Experiment 2 (I), n = 5 mice), and Experiment 3 (J), n = 6 mice) using moving-bar tasks. Format same as Figures 7I, 7J, and 7K, respectively.

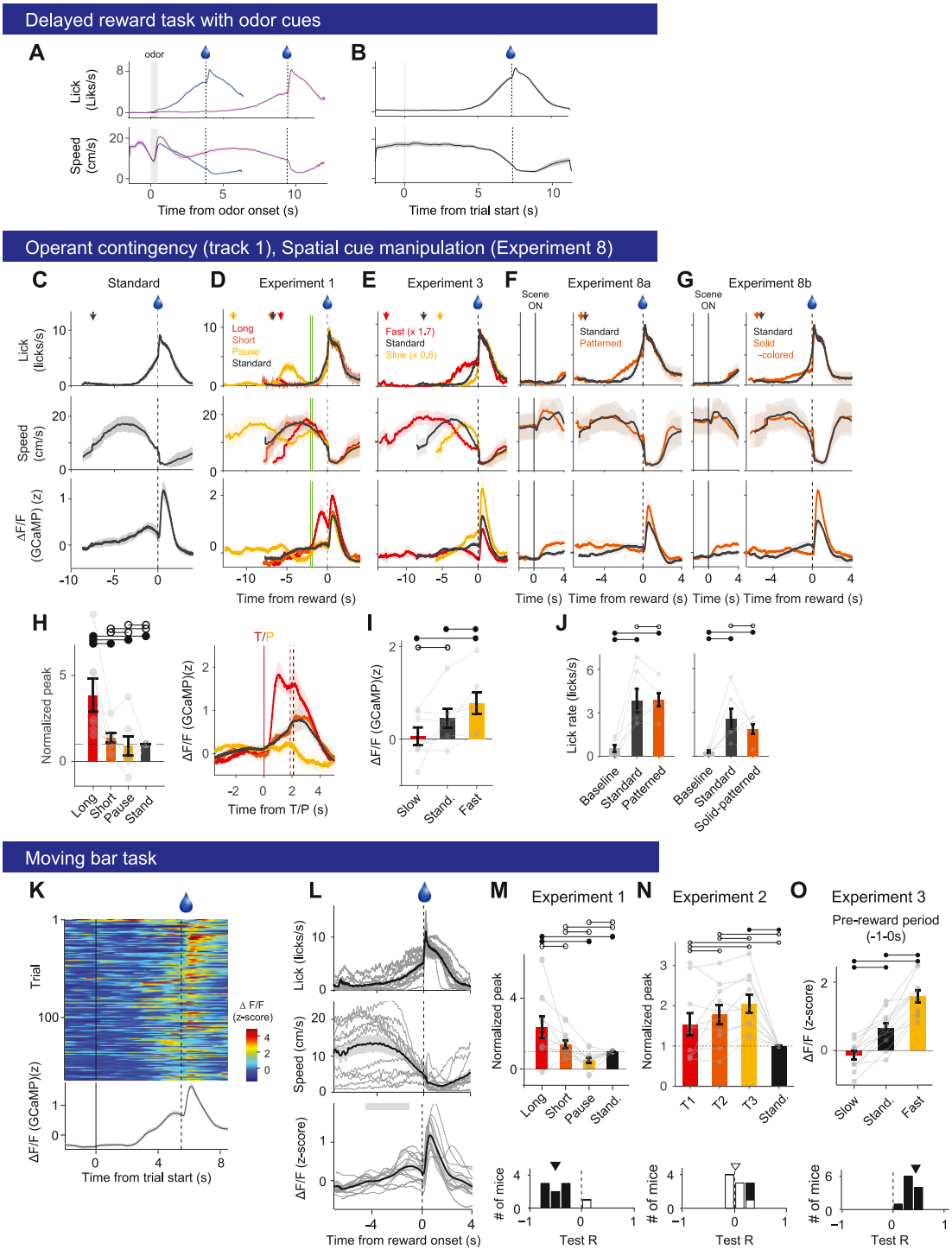



**Figure S7. Effect of Operant Contingency and Sensory Evidence on Dopamine Activity, Related to Figure 7**

(A-B) Behavioral data from delayed reward task and virtual linear track tasks. (A) Average licking (top) and running speed (bottom) in the delayed reward task with odor cues. (B) Licking (top) and running speed (bottom) in the standard virtual linear track task. (C-J) Results from the virtual linear track tasks with an operant contingency. (C) licking (top), locomotion (bottom), and dopaminergic axonal calcium signals (bottom) on the last day of training with the standard condition (n = 8 mice). Ramping *R* is smaller than passive tasks but significantly greater than zero (p = 0.02, n = 8 mice, Wilcoxon signed-rank test to compare median with zero; p = 0.48, n = 8 and 16 mice for operant and passive tasks, respectively, Wilcoxon rank-sum test). Arrow denotes the median of visual scene onsets. (D) Average lick, locomotion speed, and axonal calcium signals in Experiment 1 (n = 7 mice). Arrows denote the medians of visual scene onsets. (E) Average lick, locomotion speed, and fluorometry signals in Experiment 3 (n = 7 mice). (F) In Experiment 8a, positional cues were eliminated by using a uniform pattern on the wall in the

*(legend continued on next page)*

reduced sensory evidence condition (orange) (Figure 7E). (G) In Experiment 8b, a solid-colored scene was used not to generate optic flow in test trials (orange). (H) Quantification of the result in Experiment 1. (left) The peak responses to long teleports were significantly greater than the peak in the standard condition ($p = 0.02$, $n = 7$ mice, Wilcoxon signed-rank test) (right) dopamine activity aligned by long teleport (red), short teleport (orange), or pause (yellow) event. Response in the standard condition was aligned by the time when the animals passed the teleport destination (70 a.u.). (I) Quantification of the result in Experiment 3. Visual speed gain significantly modulated the responses before reward ($p = 0.009$, $df = 2$, $n = 7$ mice, $F = 7.05$, one-way repeated-measures ANOVA). The gain multiplier in operant version of the task (1.7) was smaller than that in the passive task (2.0) to avoid animals from aborting many trials in the slow condition. (J) Quantification of the result in Experiment 8. Lick rates during approach are significantly greater than baseline in both patterned (left) and solid-colored (right) conditions. Filled dots on top of the plot denotes significant pairwise differences, $n = 6$ mice, Wilcoxon signed-rank test). (K-O) Moving-bar experiments. (K) Dopaminergic axonal calcium signals in the VS from an example session. (L) Time courses of lick rates (top), locomotion speed (middle), and dopaminergic axon signals from individual animals (gray) as well as average across animals (black, $n = 12$ mice). Gray horizontal bar represents a time window used to compute Ramping *R*. (M-O) Statistical analysis in moving-bar tasks. (M) (top) Normalized peak responses in Experiment 1 (Figure 7I). The median of peaks in the long-distance teleport is significantly greater than 1 ($p = 0.039$, $n = 9$ mice, Wilcoxon signed-rank test). (bottom) Summary of Test *R*. The median Test *R* (black triangle) is significantly less than zero (median $r = -0.50$, $p = 0.0078$, Wilcoxon signed-rank test). (N) (top) Normalized peaks in Experiment 2 (Figure 7J). (bottom) Summary of Test *R*. The median Test *R* (open triangle) is not different from zero ($p = 0.43$, $n = 10$ mice, Wilcoxon signed-rank test). (O) (top) Results in Experiment 3 (Figure 7D). Average dopaminergic axon signals from bar movement onset to reward onset. Responses are significantly different ($n = 11$ mice, $p < 10^{-4}$, Kruskal-Wallis test). (bottom) All Test *R*s are significantly greater than zero, and median *R* is significantly greater than zero (median $r = 0.38$, $p = 0.0010$, Wilcoxon signed-rank test).