nature portfolio

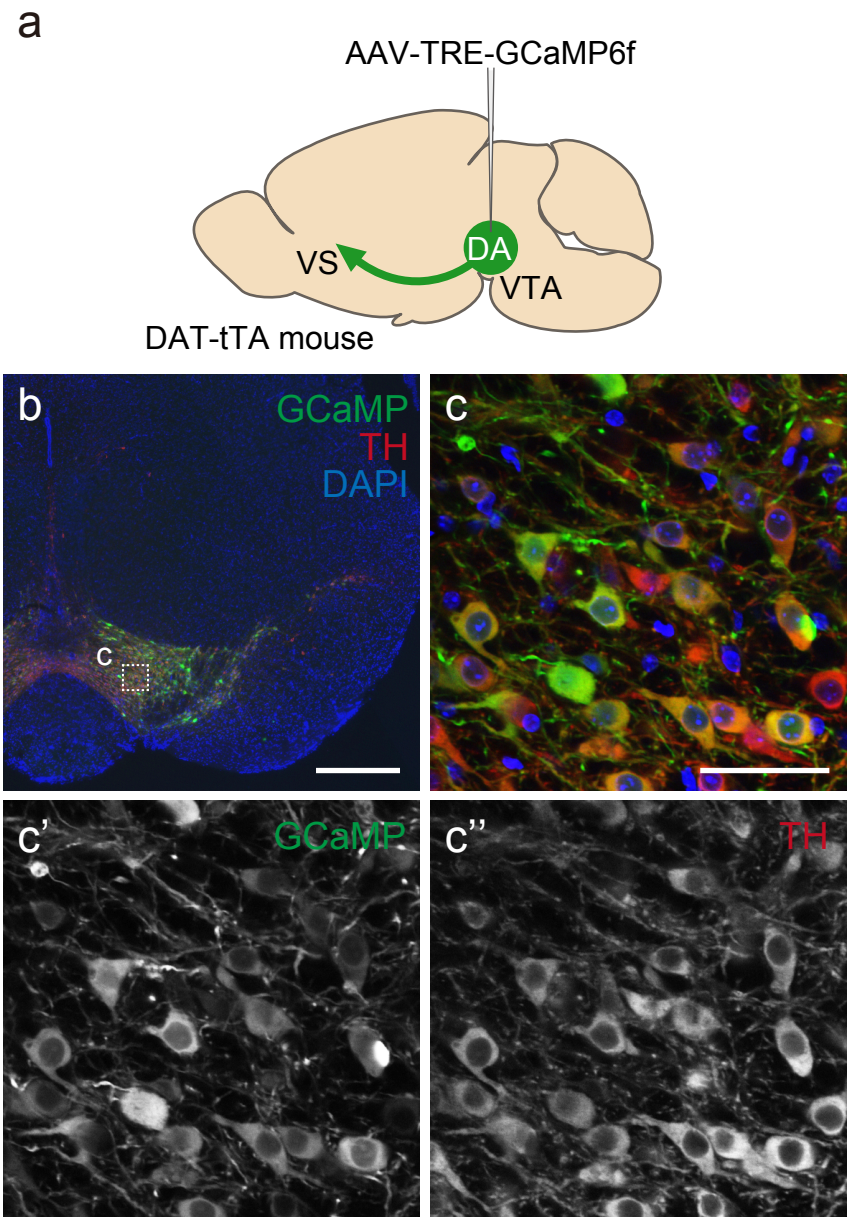
https://doi.org/10.1038/s41593-022-01109-2

Supplementary information

A gradual temporal shift of dopamine responses mirrors the progression of temporal difference error in machine learning

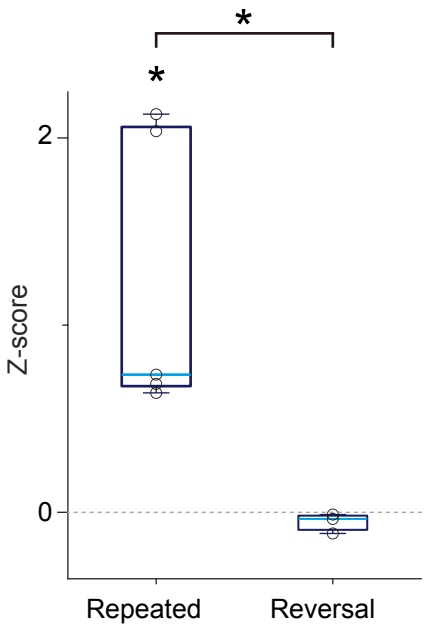
In the format provided by the authors and unedited

Supplementary Figure 1



Supplementary Figure 1. Dopamine neuron-specific GCaMP expression in DAT-tTA mice.
(**a**) tTA-dependent AAV-GCaMP (AAV5-TRE3G-GCaMP6f) was injected into the VTA in 2 animals and used for reversal learning from airpuff to reward (Figure 3d) and repeated learning (Figure 4). (**b**) A coronal section of the midbrain in DAT-tTA mouse showing expression of GCaMP (green), and dopamine neurons labeled with antibody against tyrosine hydroxylase (TH) (red). The section was counterstained with DAPI (blue). Scale bar, 500 μm. (**c**) Magnified image of the patched area in VTA in (b), showing colocalization of GCaMP signals (green) and TH immunoreactivity (red). Single channel images for GCaMP and TH immunoreactivity are shown in (c') and (c''), respectively. Scale bar, 50 μm. Number of neurons positive for both GCaMP and TH immunoreactive signals of all GCaMP positive neurons is 98.7 ± 0.5% (mean ± sem, n = 903 neurons from 3 animals) in VTA and 97.1 ± 0.8% (mean ± sem, n = 229 neurons from 3 animals) in SNc.

Supplementary Figure 2



Supplementary Figure 2. Dopamine axon response to odor in the first trial. GCaMP response to a new odor (0-1 s after odor onset) in repeated learning (Figure 3) (n = 5 animals; $t = 3.6$, $p = 0.022$ compared to the baseline; two-sided t-test) and to an odor that had previously been associated with no outcome prior to reversal learning wherein it would become associated with reward (Figure 2) (n = 3 animals; $t = -1.8$, $p = 0.22$ compared to the baseline; two-sided t-test; two-sided t-test). Responses to odor in the first trial were significantly higher in repeated learning ($t = 2.8$, $p = 0.030$; two-sided t-test). Center of boxplot showing median, edges are 25th and 75th percentile, and whiskers are most extreme data points. Center of boxplot showing median, edges are 25th and 75th percentile, and whiskers are most extreme data points. *$p<0.05$.

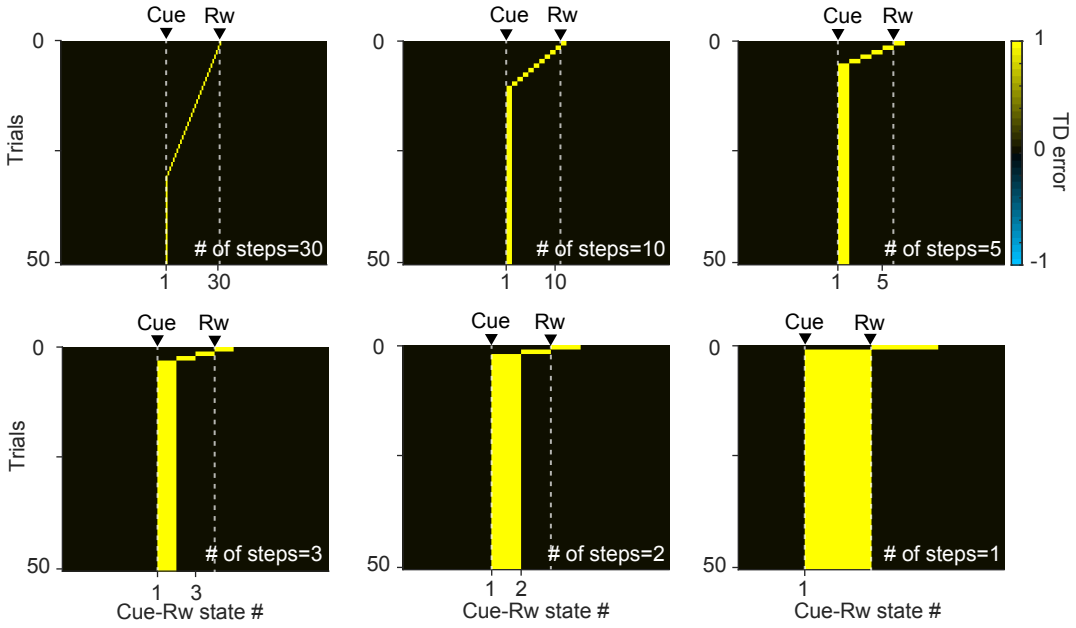
Supplementary Figure 3

a

TD(0)

learning rate (α) = 1 ($0 \leq \alpha \leq 1$)

Discounting factor (γ) = 1 ($0 \leq \gamma \leq 1$)

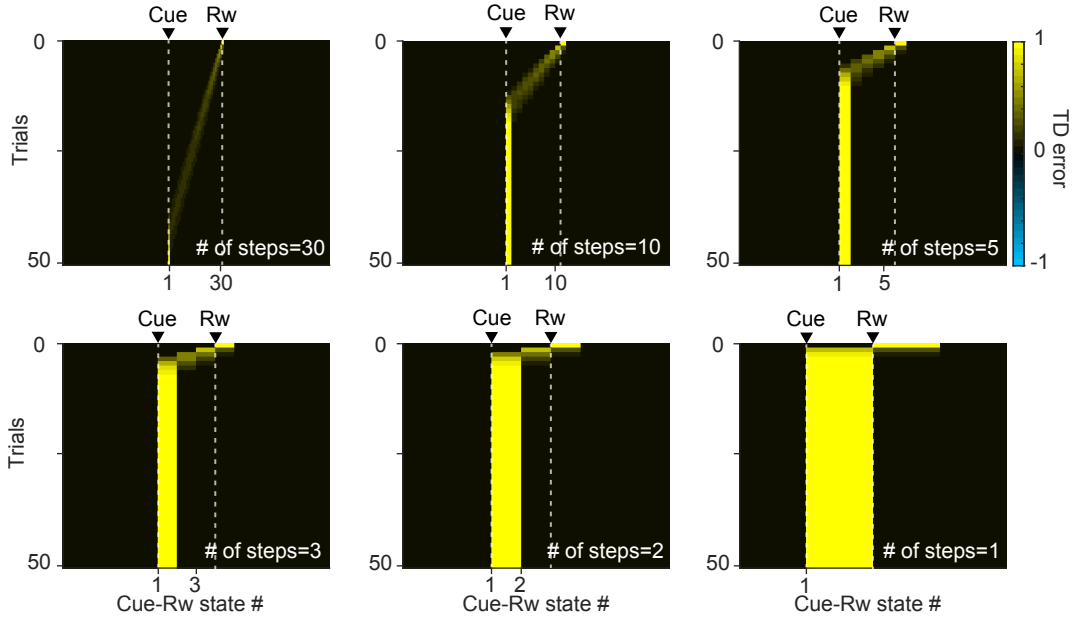


b

TD(0)

learning rate (α) = 0.7 ($0 \leq \alpha \leq 1$)

Discounting factor (γ) = 1 ($0 \leq \gamma \leq 1$)

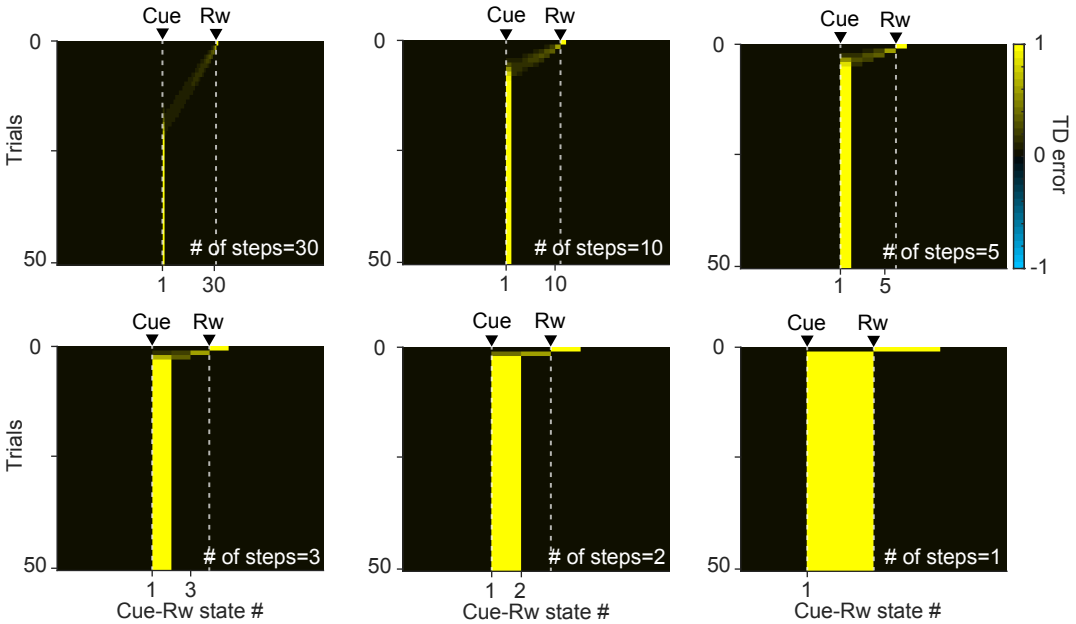


c

TD(λ) (λ=0.4)

learning rate (α) = 1 ($0 \leq \alpha \leq 1$)

Discounting factor (γ) = 1 ($0 \leq \gamma \leq 1$)



Supplementary Figure 3. TD models with different state numbers. (**a**) TD(0) model with maximum learning rate ($\alpha = 1$) with 30, 10, 5, 3, 2, and 1 state(s) between cue onset and reward onset. Gradual temporal shift is observed in all the models with >1 states between cue onset and reward. (**b**) TD(0) with moderate learning rate ($\alpha = 0.7$). Shifting activity smears temporally as the signal shifts toward the cue onset but gradual temporal shifts were observed. (**c**) TD(λ) model ($\lambda = 0.4$) with maximum learning rate ($\alpha = 1$). Shifting activity smears temporally as the signal shifts toward the cue onset but gradual temporal shifts were observed. Discounting factor $\gamma = 1$ for all models.