BRIEF COMMUNICATION
https://doi.org/10.1038/s41593-019-0574-1
nature neuroscience

Causal evidence supporting the proposal that dopamine transients function as temporal difference prediction errors

Etienne J. P Maes[1], Melissa J. Sharpe[2], Alexandra A. Usypchuk[1], Megan Lozzi[1], Chun Yun Chang[3], Matthew P. H. Gardner[3], Geoffrey Schoenbaum[3,4,5*] and Mihaela D. Iordanova[1*]

Reward-evoked dopamine transients are well established as prediction errors. However, the central tenet of temporal difference accounts—that similar transients evoked by reward-predictive cues also function as errors—remains untested. In the present communication we addressed this by showing that optogenetically shunting dopamine activity at the start of a reward-predicting cue prevents second-order conditioning without affecting blocking. These results indicate that cue-evoked transients function as temporal-difference prediction errors rather than reward predictions.

One of the most fundamental questions in neuroscience concerns how associative learning is implemented in the brain. Key to most implementations is the concept of a prediction error—a teaching signal that supports learning when reality fails to match predictions[1]. The greater the error, the greater the learning. In computational accounts, these errors are calculated by the method of temporal difference[2,3], in which time (t) is divided into states, each containing a value prediction (V) derived from past experience that is the basis of a rolling prediction error. This error (δ) is the difference between successive states. The most famous of these is temporal difference reinforcement learning[3], the prediction error of which

$$\delta(t) = V(t) - V(t-1)$$

has been mapped on to millisecond-resolution changes in dopamine neuron firing[4].

Although this mapping has been one of the signature success stories of modern neuroscience, one pillar of this account that has not been well-tested is that the transient increase in firing evoked by a reward-predicting cue is a temporal difference error, propagated back from the reward and functioning to support learning about predictors of that cue. Evidence for true, gradual back-propagation of this signal is sparse, as is evidence that it exhibits signature features that define the error at the time of reward, such as suppression on omission of the cue when it has been predicted by an earlier cue, and transfer back to such earlier predictors (in the absence of the primary reward itself). Furthermore, there is little or no causal evidence that cue-evoked dopamine serves as an error signal to support learning. Indeed, the cue-evoked signal is often described as if it encodes the cue's importance or value derived from its prediction of future reward. Such language is imprecise, leading at best to confusion about the theorized unitary function of the dopamine transient and at worst to a true dichotomization of the function of cue- versus reward-evoked activity. This situation is especially curious, because the appearance of the dopamine transient in response to reward-predictive cues is a lynchpin of the argument that the dopamine neurons signal a temporal difference error[1].

A logical way to address this question is to test, using second-order conditioning[5], whether optogenetic blockade or shunting of dopamine activity at the start of a reward-predictive cue prevents learning about this cue in the same way that optogenetic shunting of dopamine activity at reward delivery prevents learning about reward[6]. If this signal is a temporal difference error, $\delta(t_{cue})$ in the terms of the above equation, then blocking it will prevent such learning (Fig. 1, and see Extended Data Fig. 1), which shows an experimental design for second-order conditioning and computational modeling of the effect of eliminating $\delta(t_{cue})$. However, although this seems at first like a conclusive experiment, it is not, because the same effect is obtained by eliminating the cue's importance or ability to predict reward for the purposes of calculating the prediction error (Fig. 1, and see Extended Data Fig. 1). This occurs because the ability of the cue to predict reward is the source of $V(t_{cue})$, which is the basis of the cue-evoked prediction error. If shunting the transient eliminates the cue-evoked prediction, then it could also eliminate the cue-evoked prediction error. Consequently, the disruption of second-order conditioning by shunting of the dopamine transient would show that this signal is necessary for learning, but it would not distinguish whether this is because it is a prediction error or a reward prediction.

This confound can be resolved by combining the above experiment with an assessment of the effects of the same manipulation (ideally in the same subjects) on blocking[7]. Blocking refers to the ability of a cue to prevent or block other cues from becoming associated with the predicted reward; blocking is thought to reflect the reduction in prediction error at the time of the reward, $\delta(t_{rew})$, caused by the cue's contribution to the reward prediction in the reward state ($V(t_{rew}-1)$). If the cue-evoked dopamine transient is carrying the cue's reward prediction, then optogenetically shunting it should diminish or prevent blocking, because in the absence of the cue's reward prediction the reward would still evoke a prediction error (Fig. 1, and see Extended Data Fig. 1), which shows an experimental design for blocking and computational modeling of the effect of eliminating $V(t_{rew}-1)$. On the other hand, if the

[1]Department of Psychology/Centre for Studies in Behavioural Neurobiology, Concordia University, Montreal, Quebec, Canada. [2]Department of Psychology, University of California, Los Angeles, CA, USA. [3]Intramural Research Program, National Institute on Drug Abuse, Baltimore, MD, USA. [4]Departments of Anatomy & Neurobiology and Psychiatry, University of Maryland School of Medicine, Baltimore, MD, USA. [5]Solomon H. Snyder Department of Neuroscience, The Johns Hopkins University, Baltimore, MD, USA. *e-mail: geoffrey.schoenbaum@nih.gov; mihaela.iordanova@concordia.ca

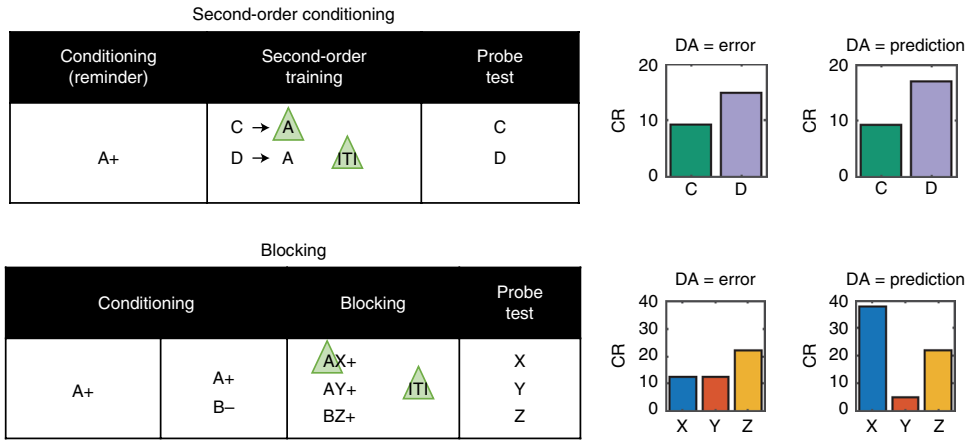


Fig. 1 | Modeling results. Experimental designs for second-order conditioning (top row) and blocking (bottom row), along with bar graphs modeling the predicted results of shunting the dopamine transient at the start of the reward-predictive cue, A, in each procedure. Green triangles indicate light delivery to shunt dopamine transients at the start of the reward-predicting cue (i.e., A) or during the intertrial interval (ITI) in TH-Cre rats expressing halorhodopsin in VTA neurons. The left column of bar graphs shows modeled results under the hypothesis that the cue-evoked dopamine transient signals a prediction error; the right column shows them under the hypothesis that it signals a reward prediction. Elimination of either signal would impair second-order conditioning (top graphs, C versus D), but only elimination of a prediction would affect blocking (bottom graphs, X versus Y and Z). Note the output of the classic temporal difference reinforcement learning model was converted from V to CR to better reflect the behavioral output actually measured in our experiments (see Methods for details). See Extended Data Fig. 1 for modeling of behavior in the full experiments, culminating in these displays.

cue-evoked dopamine signal reflects only the actual prediction error occurring at the start of the cue, $\delta(t_{cue})$, then its removal should have no impact on the blockade of the prediction error evoked by the reward and, thus, no impact on blocking (Fig. 1, and see Extended Data Fig. 1).

Armed with these contrasting, computationally validated predictions, we conducted a within-subject version of the designs (Fig. 1, and see Extended Data Fig. 1). Sixteen Long–Evans transgenic rats expressing Cre recombinase under control of the tyrosine hydroxylase promoter (TH-Cre$^{+/-}$) served as subjects. Four weeks before the start of testing, the rats underwent surgery to infuse a Cre-dependent viral vector carrying halorhodopsin (NpHR) (AAV5-EF1α-DIO, eNpHR3.0-eYFP) into the ventral tegmental area (VTA) bilaterally and to implant optical fibers targeting this region (see Extended Data Fig. 2). Rats were food restricted immediately before the start of testing and then trained to associate a visual cue, A, with a reward. The food port approach (percentage time spent) was taken as a measure of the level of conditioning. The blocking experiment was carried out in the same rats before the second-order conditioning (SOC) experiment to ensure that any effect of non-reinforcement during SOC training would not disrupt blocking. However, to align with the logic of our modeling, we present the SOC data here first.

Second-order training consisted of two sessions in which A was presented without reward, preceded on each trial by one of two 10-s novel auditory cues, C or D. On C→A trials, a continuous laser light (532 nm, 18–20 mW output) was delivered into the VTA for 2.5 s, starting 0.5 s before the onset of A to disrupt any dopamine transient normally occurring at the start of the reward-predicting cue, in this case A. We model this as shunting, because we have found that similar duration patterns of inhibition disrupt learning from positive errors without inducing aversion or learning from negative errors[6,8,9]. On D→A trials, the same 2.5-s light pattern was delivered during the intertrial interval at a random time point 120–180 s after termination of A. After this training, rats underwent probe testing, in which C and D were presented alone and without reward (Fig. 2, with supporting statistics described in the legend). There was no difference in responding on C→A versus D→A trials, indicating that the optogenetic manipulation did not deter responding

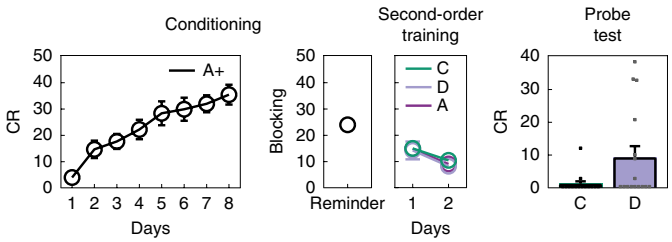


Fig. 2 | The cue-evoked dopamine transient is necessary for SOC. Behavioral responding (mean + s.e.m., $n = 16$ rats) during A increased during conditioning (ANOVA: $F_{1,15} = 65.6$, $P < 0.005$) and after blocking during reminder training (ANOVA: $F_{1,15} = 38.0$, $P < 0.001$). Responding to C (that is, C→A trials) and D (that is, D→A trials) did not differ (Friedman's ANOVA: for day 1 $\chi^2(1,31) = 0.6$, $P = 0.439$; for day 2 $\chi^2(1,31) = 0.29$, $P = 0.593$) during second-order training when shunting of VTA transients took place at the start of the reward-predictive cue, A (as illustrated in Fig. 1). Wilcoxon's signed-rank test was used to analyze the data on the probe test. Responding to C was lower compared with D ($z = 1.9$, $P = 0.03$, effect size $r = 0.34$), showing that inhibition of the VTA DA signal at the start of A prevented A from supporting SOC to C, whereas identical inhibition during the ITI left learning to D intact. CR is the percentage time spent in the magazine during the last 5 s of the cue.

during second-order training. However, in the probe test, responding to C was notably lower than responding to D, indicating that light delivery at the start of A prevented second-order conditioning of C. Identical results were obtained in a separate experimental group that underwent the same training without the prior experience with blocking, described below (see Extended Data Fig. 3), whereas the yellow fluorescent protein eYFP controls, which underwent the same training after being transfected with a virus lacking NpHR2.0, showed high levels of responding to both C and D (see Extended Data Fig. 2).

Blocking consisted of four sessions in which A was presented in a compound with two novel 10-s auditory cues, X and Y, followed by a reward. On AX trials, continuous laser light (532 nm, 18–20 mW

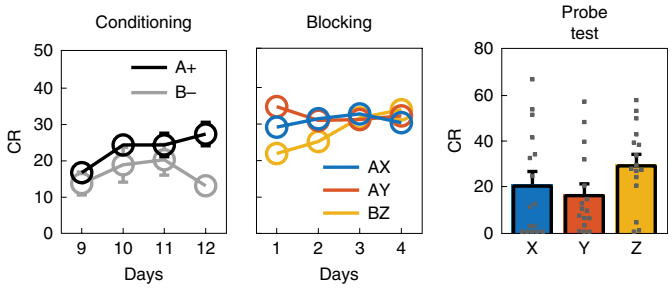


Fig. 3 | The cue-evoked dopamine transient is not necessary for blocking. Behavioral responding (mean + s.e.m., $n = 16$ rats) during conditioning was greater for the reinforced cue, A, compared with the non-reinforced cue, B (ANOVA: $F_{1,15} = 14.9$, $P = 0.002$; see also Methods). During blocking, responding to the control compound (BZ) was lower compared with that seen for the blocking compounds (AX and AY) on the first day of training for NpHR (ANOVA: $F_{1,15} = 12.9$, $P = 0.003$), but similar on subsequent days (ANOVA: max. $F_{1,15} = 2.7$, $P = 0.122$); shunting of the VTA DA transient took place at the start of the reward-predicting cue, A (as illustrated in Fig. 1), yet responding to AX and AY was similar on each day (ANOVA: max. $F_{1,15} = 1.3$, $P = 0.272$). Wilcoxon's signed-rank test was used to analyze the data on the probe test. The rats showed a blocking effect: there was higher level of responding to the control cue (Z) compared with the blocked cues (X and Y) ($z = 2.22$, $P = 0.01$, $r = 0.39$) on the first trial pooled across both probe tests. There was no effect of VTA DA inhibition on blocking because responding to the blocked cues (X versus Y) did not differ ($z = 0.05$, $P = 0.48$, $r = 0.009$). CR is the percentage time spent in the magazine during the last 5 s of the cue.

output) was delivered into the VTA for 2.5 s, starting 0.5 s before the onset of A; on AY trials, the same 2.5-s light pattern was delivered during the intertrial interval at a random time point, starting 120–180 s after termination of the compound. In addition, as a positive control for learning about a compound cue, the rats also received presentations of two additional cues, B and Z, followed by the same reward. B was a visual cue, which was presented four times without reward on each of the last 4 d of conditioning (see Extended Data Fig. 2). Z was a third novel auditory cue. After this training, rats underwent probe testing, in which X, Y and Z were presented alone and without a reward (Fig. 3, with supporting statistics described in the legend). During blocking, responding to BZ was lower at the start of training but reached that of the AX and AY compounds by the end of training. There were no differences in responding on AX versus AY trials across blocking, indicating that the optogenetic manipulation did not deter responding. In the probe test, responding to the positive control cue, Z, was notably higher than responding to two blocked cues, X and Y, indicating that pretraining of A blocked learning for these two cues. Furthermore, there was no difference in responding to X and Y, indicating that light delivery at the start of A on AX trials had no effect on blocking. Identical results were obtained in eYFP control rats (see Extended Data Fig. 2).

These data provide clear and concise evidence that transient increases in the firing of dopamine neurons at the start of reward-predictive cues function as prediction errors to support associative learning, in much the same way that reward-evoked changes have been shown to do. As noted in our introduction, such evidence is important because the proposal that the cue-evoked dopamine transient is a prediction error is the lynchpin of the hypothesis that dopaminergic error signals integrate information about future events, thereby providing a temporal difference error. The finding that dopamine neuron activity at the start of a reward-predicting cue is necessary for SOC but not blocking provides strong support for this idea, while at the same time ruling out alternative proposals that this signal—at least at the level of the spiking of dopamine neurons—reflects the actual associative importance of the cue with respect to predicting reward (but see ref. [10]). Importantly, our findings are agnostic with regard to the nature of the information in the temporal difference signal or the specific type of learning that it supports. This is a noteworthy caveat, because temporal difference errors can be limited to representing information about value[2,3] or they can be construed more broadly as representing errors in predicting other value-neutral information[2,11]. Recent studies using sensory preconditioning and reinforcer devaluation provide evidence that dopamine transients can support learning orthogonal to value in line with the latter account[9,11,12]. In the present communication, we used SOC, which has been proposed to rely on an associative structure that bypasses the representation of the outcome and links a stimulus and a response[13]. Sensory preconditioning, by contrast, is supported by the association of two neutral stimuli, leaving no opportunity for direct links with a reward-based response. The fact that dopamine transients in the VTA are now causally implicated in supporting both forms of learning supports a much broader role for these signals in driving associative learning than is envisioned by current dogma, one in which the content of the learning supported is determined by the learning conditions.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41593-019-0574-1.

Received: 9 January 2019; Accepted: 9 December 2019;
Published online: 20 January 2020

References

1. Glimcher, P. W. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proc. Natl Acad. Sci. USA* **108**, 15647–15654 (2011).
2. Dayan, P. Improving generalization for temporal difference learning: the successor representation. *Neural Comput.* **5**, 613–624 (1993).
3. Sutton, R. S. Learning to predict by the method of temporal difference. *Machine Learn.* **3**, 9–44 (1988).
4. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate for prediction and reward. *Science* **275**, 1593–1599 (1997).
5. Rizley, R. C. & Rescorla, R. A. Associations in second-order conditioning and sensory preconditioning. *J. Compar. Physiol. Psychol.* **81**, 1–11 (1972).
6. Chang, C. Y., Gardner, M., Di Tillio, M. G. & Schoenbaum, G. Optogenetic blockade of dopamine transients prevents learning induced by changes in reward features. *Curr. Biol.* **27**, 3480–3486 (2017).
7. Kamin, L. J. Aversive stimulation. In *Miami Symposium on the Prediction of Behavior*, 1967 (ed. M.R. Jones) 9–31 (Univ. Miami Press, 1968).
8. Chang, C. Y., Gardner, M. P. H., Conroy, J. S., Whitaker, L. R. & Schoenbaum, G. Brief, but not prolonged, pauses in the firing of midbrain dopamine neurons are sufficient to produce a conditioned inhibitor. *J. Neurosci.* **38**, 8822–8830 (2018).
9. Sharpe, M. J. et al. Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nat. Neurosci.* **20**, 735–742 (2017).
10. Kim H. R. et al. A unified framework for dopamine signals across timescales. Preprint at *bioRxiv* https://doi.org/10.1101/803437 (2019).
11. Gardner, M. P. H., Schoenbaum, G. & Gershman, S. J. Rethinking dopamine as generalized prediction error. *Proc. R. Soc. B* **285**, https://doi.org/10.1098/rspb.2018.1645 (2018).
12. Keiflin, R., Pribut, H. J., Shah, N. B. & Janak, P. H. Ventral tegmental dopamine neurons participate in reward identity predictions. *Curr. Biol.* **29**, 93–103.E3 (2019).
13. Nairne, J. S. & Rescorla, R. A. 2nd-order conditioning with diffuse auditory reinforcers in the pigeon. *Learn. Motiv.* **12**, 65–91 (1981).

Methods

Modeling. Simulations of the behavioral designs were run using a one-step temporal difference learning algorithm, TD(0)[14]. This algorithm was used to estimate the value of different states of the behavioral paradigm with states being determined by the stimuli present at any particular time. Linear function approximation was used to estimate the value, V, of a given state, s_t, by the features present during that state according to:

$$\hat{V}(s_t) \approx \sum_j \mathbf{w}_j \mathbf{x}_j(s_t)$$

where j is indexed through all possible components of the feature vector $\mathbf{x}$ and corresponding weight vector $\mathbf{w}$. The feature vector is considered to be the set of possible observed stimuli such that, if stimulus j is present during state s at time t, then $\mathbf{x}_j(s_t)=1$, and 0 otherwise. The weights are adjusted over time to give the best approximation of the value of each state given the current set of stimuli. Weights, $\mathbf{w}_j$, corresponding to each feature, $\mathbf{x}_j$, are updated at each time step according to the temporal difference error rule:

$$\delta_t = \left[r_t + \gamma \hat{V}(s_t) - \hat{V}(s_{t-1})\right]$$

under linear value function approximation where γ is the temporal discounting factor. The weights are updated as:

$$\Delta \mathbf{w}_j = \alpha \mathbf{x}_j(s_t)\delta_t$$

in which the scalar α is the learning rate. The linear value approximation reduces the size of the possible state space by generalizing states based on the features present. This approximation results in the calculation of the total expected value of a state as the sum of the expected value of each stimulus element present in the current state, a computation that is consistent with a global prediction error as stipulated by the Rescorla–Wagner model[15].

Modeling of optogenetic manipulation of midbrain dopamine activity. Optogenetic inhibition of dopaminergic neurons was modeled two different ways to align with the different hypotheses of dopamine function described earlier (see Main).

Model 1: Dopamine transients correspond to temporal difference errors. For this model, inhibition of dopaminergic activity disrupts solely the error signal[11]

$$\delta_t = \eta_t\left[r_t + \gamma \hat{V}(s_t) - \hat{V}(s_{t-1})\right] \quad \eta = \begin{cases} 0\ \text{“}laseron\text{”} \\ 1\ \text{“}laseroff\text{”} \end{cases}$$

where η is a binary value determining whether the inhibition was present or not during state s_t.

Model 2: Dopamine transients correspond to expected value. In this case, the dopaminergic inhibition disrupts the future expected value during the current state and, as this becomes the prior expected value in the next state, the inhibition disrupts this as well:

$$\delta_t = \left[r_t + \gamma\eta_t \hat{V}(s_t) - \eta_{t-1}\hat{V}(s_{t-1})\right] \quad \eta = \begin{cases} 0\ \text{“}laseron\text{”} \\ 1\ \text{“}laseroff\text{”} \end{cases}$$

where η again determines whether the inhibition was present.

Model parameterization. Generalization of value across stimuli was modeled by setting the initial weights, $\mathbf{w}_j$, of a stimulus to 0.7 for stimuli of the same modality and 0.2 for stimuli of different modalities.

Conditioned responding to the food cup, CR, at each state was modeled using a logistic function:

$$\mathrm{CR}(s_t) = \frac{c}{1 + e^{-b(V(s_t)-a)}}$$

in which the parameters were determined based on empirical estimates of the maximal responding, c, the baseline responding, a, as well as the steepness of the learning curve, b. These were set as 55, 0.4 and 3, respectively, for all simulations. Reduced responding to the food cup while rats were attached to the patch cables was modeled as a reduction in the maximal responding to 40.

All simulations were performed with $\alpha=0.05$ and $\gamma=0.95$. To ensure that the order of cue presentations did not affect the findings, cue presentations during each stage of conditioning were pseudo-randomized and the results of the simulations were averaged over 100 repetitions of the model. Simulations were performed using custom-written functions in MATLAB (Mathworks), which are available in the Supplementary Software and are posted on Github (https://github.com/mphgardner/Basic_Pavlovian_TDRL/tree/Maes_2018).

Subjects. A total of 45 experimentally naive Long–Evans transgenic rats expressing Cre recombinase under control of the tyrosine hydroxylase promoter (TH-Cre$^{+/-}$) were used in the experiments reported in the present communication. The rats were approximately 3 months of age at the start of the experiment. Of those rats 16 were bred in-house at the National Institute on Drug Abuse (NIDA; male: $n=9$, 390–587 g; female: $n=7$, 302–370 g) and 14 rats were bred in-house at Concordia University (male: $n=5$, 382–515 g; female: $n=7$, 247–289 g) and infused with a viral vector carrying NpHR (see below); the remaining 15 rats were bred in-house at NIDA (male: $n=8$, 450–630 g; female: $n=7$, 250–330 g) and infused with a control viral vector (eYFP only, see below). Sample sizes were chosen based on published work[6,8,9]. Four rats were excluded from the Concordia-bred cohort due to lack of virus expression ($n=1$), failure to consume the pellets during conditioning ($n=1$), or failure to receive stimulation due to broken ferrules ($n=1$) or cables ($n=1$). Four rats were excluded from the eYFP group due to lack of virus expression ($n=2$), failure to receive stimulation due to a broken cable ($n=1$) or a significant outlier result according to Grubb's test ($n=1$; $Zc=2.55$, $Z=2.8$, $P<0.05$; https://www.graphpad.com/quickcalcs/Grubbs1.cfm). There were no effects of sex across the different phases in our study (NpHR: max. $F_{1,14}=2.1$, $P=0.17$; eYFP: max. $F_{1,14}=4.0$, $P=0.08$). The rats were implanted with bilateral optical fibers in the VTA at approximately 4 months of age. Please refer to the Life Sciences Reporting Summary for additional information.

Surgical procedures. Surgical procedures have been described elsewhere[8,9]. Rats were infused bilaterally with 1.2 µl AAV5-EF1α-DIO-eNpHR3.0-eYFP or AAV5-EF1α-DIO-eYFP into the VTA at the following coordinates relative to bregma: AP: −5.3 mm; ML: ±0.7 mm; DV: −6.55 and −7.7 mm (females) or −7.0 and −8.2 mm (males). The viral vector was obtained from the Vector Core at the University of North Carolina at Chapel Hill. During surgery, ferrules carrying optical fibers were implanted bilaterally (200 µm diameter, Precision Fiber Products) at the following coordinates relative to bregma: AP: −5.3 mm; ML: ±2.61 mm, and DV: −7.05 mm (female) or −7.55 mm (male) at an angle of 15° pointed toward the midline.

Apparatus. The within-subject NpHR and eYFP experiments were conducted using eight behavioral chambers (Coulbourn Instruments), which were individually housed in light- and sound-attenuating cabinets. The replication of the SOC study was conducted using eight behavioral chambers (Med-Associates) which were individually housed in light-attenuating, custom-made cabinets. Each chamber was equipped with a pellet dispenser that delivered 45-mg sucrose pellets into a recessed magazine when activated. Access to the magazine was detected by means of infrared detectors mounted across the opening of the recess. Two light panels (NIDA: differently shaped; Concordia: identical) were located on the right-hand wall of the chamber above and on either side of the magazine. At NIDA the chambers contained a speaker housed within the chambers whereas, at Concordia, the chambers contained two speakers located outside the testing chamber but inside the housing cabinet. The speakers were connected to a custom-built Arduino device containing wave files of the stimuli used (NIDA: 5-Hz clicker, white noise, tone, siren, chime; Concordia: white noise, 4-Hz clicker). The stimulus intensity was 72–74 dB. A computer equipped with Coulbourn Instruments GS3 or Med-Associates Med-IVR software controlled the equipment and recorded the responses.

Housing. Rats were housed singly and maintained on a 12-h light–dark cycle. Behavioral experiments took place during the light cycle at NIDA and during the dark cycle at Concordia. Rats had free access to food and water unless undergoing behavioral testing, during which they received sufficient chow to maintain them at ~85% of their free-feeding body weight. All experimental procedures conducted at the NIDA–IRP were in accordance with the Institutional Animal Care and Use Committee of the US National Institutes of Health (NIH) guidelines, and those conducted at Concordia University were in accordance with the approval granted by the Canadian Council on Animal Care and the Concordia University Animal Care Committee.

General behavioral procedures. Trials consisted of 13-s visual and 10-s auditory cues as described below; visual cues were 3 s longer and their onset was 3 s before auditory cue onset in the blocking part of the study; in the SOC part of the study, the auditory cues preceded the visual cues with auditory cue offset coinciding with visual cue onset. In the SOC replication study done at Concordia University, the visual and auditory cues were 10 s long. These cue arrangements allowed for optogenetic manipulation of the dopamine transient at the start of visual cue A without any interference with the processing of other cues. Trial types were interleaved in miniblocks, with the specific order unique to each rat and counterbalanced across groups. Intertrial intervals varied around a 6 min mean (range 4–8 min). All rats were trained between 10am and 8pm. Five auditory stimuli (tone, clicker, white noise for X, Y and Z in blocking; chime and siren (NIDA) or white noise and clicker (Concordia) for C and D in SOC) and two visual stimuli (flashing light and steady light for A and B) were used. The stimuli were counterbalanced across rats within each modality, and the reward used throughout consisted of two 45-mg sucrose pellets (NIDA: no flavor; Concordia: chocolate flavored; 5TUT, TestDiet).

Training for the blocking/SOC, within-subject experiment consisted of six phases: conditioning, blocking, blocking probe test, reminder training, SOC and

second-order probe test. These are described below. Training for the second-order replication experiment follows.

Conditioning. Conditioning took place across 12 d (8 d untethered, 4 d tethered) and each day consisted of 14 presentations of A→2US, where a 13-s presentation of A was immediately followed by two 45-mg sucrose pellets (5TUT, TestDiet). Toward the end of conditioning (on tethered days 9–12), the rats also received four trials per day of non-reinforced presentations of B. This was done to reduce unconditioned orienting to the novel visual stimulus that would detract from learning on the first few trials of the compound stimulus[16]. Conditioning data were normally distributed (NpHR: for A, $P=0.739$; for B, $P=0.084$; eYFP: for A, $P=0.984$; for B, $P=0.118$). Responding (see Fig. 2 and also Extended Data Fig. 2: conditioning) to A increased across the first 8 d of conditioning (NpHR: $F_{1,15}=65.6$, $P<0.005$; eYFP: $F_{1,10}=94.4$, $P<0.005$). During the subsequent 4 d of discrimination training, responding was higher for A compared with B (see Fig. 3 and also Extended Data Fig. 2: conditioning NpHR: $F_{1,15}=14.9$, $P=0.002$; eYFP: $F_{1,10}=16.3$, $P<0.001$), but it remained stable (NpHR: $F_{1,15}=3.0$, $P=0.106$; eYFP: $F_{1,10}=3.1$, $P=0.099$) and there was no interaction (NpHR: $F_{1,15}=5.2$, $P=0.038$; eYFP: $F_{1,10}=6.6$, $P<0.021$).

Blocking. After conditioning, all rats received 4 d of compound conditioning, that is blocking. Blocking followed the initial conditioning phase because it was paramount that the reinforced cue had not been experienced in the absence of reinforcement (as in SOC) because this could compromise its effectiveness to block learning. During this phase two compounds, consisting of the pretrained cue A and a novel auditory cue, X or Y, and a third compound, consisting of the pre-exposed cue B and a novel auditory cue Z, were presented. Each compound received six reinforced trials with the same reward (AX→2US; AY→2US; BZ→2US). This yielded two blocking compounds, AX and AY, and a control compound, BZ. The presentation of the blocking and blocked cues were offset (see also ref. [17]) such that the 13-s visual cues began 3 s before onset of the 10-s auditory cues. On AX trials, continuous laser light (532 nm, 18–20 mW output, Shanghai Laser & Optics Century) was delivered into the VTA for 2.5 s, starting 0.5 s before the onset of A; on AY trials, the same light pattern was delivered during the intertrial interval, 120–180 s after termination of the compound. Data during the blocking phase were normally distributed (NpHR: for AX, $P=0.560$; for AY, $P=0.802$; for BZ, $P=0.568$; eYFP: for AX, $P=0.555$; for AY, $P=0.675$; for BZ, $P=0.875$). During blocking (see Fig. 3 and also Extended Data Fig. 2: blocking), responding to the control compound (BZ) was lower compared with that seen to the blocking compounds (AX and AY) on the first day of training for NpHR ($F_{1,15}=12.9$, $P=0.003$) but not for eYFP ($F_{1,10}=2.0$, $P=0.19$), but was similar on subsequent days (NpHR: day 2, $F_{1,15}=2.7$, $P=0.122$; day 3, $F<1$, $P=0.934$; day 4, $F<1$, $P=0.422$; eYFP: day 2, $F_{1,15}=1.1$, $P=0.328$; day 3, $F_{1,10}=2.7$, $P=0.134$; day 4, $F<1$, $P=0.950$). The response to the blocking compounds (AX and AY) did not differ across this phase of training for NpHR (day 1, $F_{1,15}=1.3$, $P=0.272$; day 2, $F<1$, $P=0.918$; day 3, $F<1$, $P=0.703$; day 4, $F<1$, $P=0.593$) or for eYFP except for day 1 (day 1 $F_{1,10}=5.8$, $P=0.037$; day 2, $F<1$, $P=0.702$; day 3, $F_{1,10}=1.3$, $P=0.281$; day 4, $F_{1,10}=2$, $P=0.185$). The lack of differences between AX and AY provides evidence that shunting DA firing during the start of A did not disrupt the processing of A.

Blocking probe test. To confirm learning and determine the effect of inhibition of TH$^+$ neurons in the VTA, rats received a probe test in which each of the auditory cues (X, Y and Z) was presented 4 times alone and without reward for a total of 12 trials. Rats received the same probe test again 2 d later, which allowed for behavioral recovery. The two test sessions were collapsed. Analyses focused on the pooled data from the initial trial in each test. The first trial eliminates any within-session effects of non-reinforcement, which can mask behavioral differences (see also ref. [18]). Data from the test were not all normally distributed (NpHR: for X, $P=0.017$; for Y, $P=0.012$; for Z, $P=0.441$; eYFP: for X, $P=0.022$; for Y, $P=0.244$; for Z, $P=0.854$), so Wilcoxon's signed rank test was used to analyze differences between the conditions (see the legend for Fig. 3). On the test, the rats showed a blocking effect: there was a higher level of responding to the control cue (Z) compared with the blocked cues (X and Y) (NpHR: Fig. 2, $z=2.22$, $P=0.01$, $r=0.39$; eYFP: Extended Data Fig. 2, $z=1.96$, $P=0.03$, $r=0.42$). There was no effect of VTA DA inhibition (that is, NpHR condition) on blocking because responding to the blocked cues (X versus Y) did not differ ($z=0.05$, $P=0.48$, $r=0.009$). There was also no effect of VTA light stimulation (eYFP condition) because responding to the blocked cues (X versus Y) did not differ ($z=0.97$, $P=0.17$, $r=0.206$). We also compared the pooled data from all trials for both the NpHR and eYFP groups of rats together. There was no effect of group ($F<1$, $P=0.573$), there was a difference between the control (Z) compared with the blocked cues (X and Y; $F_{4,2}=7.6$, $P=0.01$), but this difference did not interact with the group ($F<1$, $P=0.561$), there was no difference between the blocked cues ($F<1$, $P=0.850$) and no interaction between the blocked cues with the group ($F<1$, $P=0.430$).

Reminder training. Before the start of SOC, all rats received a single reminder session for A, which consisted of re-training of the A→2US contingency across 14 trials as described above. This was done to offset any effects of probe testing without reward at the end of blocking. Magazine responding to the retrained cue A (see Fig. 2 and also Extended Data Fig. 2) was normally distributed (NpHR: $P=0.432$; eYFP: $P=0.348$) and was found to increase across trials (NpHR: $F_{1,15}=38.0$, $P<0.001$; eYFP: $F_{1,15}=25.6$, $P<0.001$).

Second-order conditioning (SOC). After re-training, rats received two sessions of SOC consisting of six presentations of C and six presentations of D, each paired with A (C→A; D→A). No rewards were delivered during this phase. On C→A trials, continuous laser light (532 nm, 18–20 mW output, Shanghai Laser & Optics Century) was delivered into the VTA at the start of A, in the same manner as that used in blocking; on D→A trials, the same light pattern was delivered during the intertrial interval, 120–180 s after termination of A. Responding during this phase was generally not normally distributed (NpHR: for C, $P=0.024$; for D, $P=0.009$; for A(C), $P=0.069$; for A(D), $P=0.041$; eYFP: for C, $P=0.019$; for D, $P=0.213$; for A(C), $P<0.001$; for A(D), $p=0.026$). Therefore, Friedman's ANOVA was used to analyze responding during this phase. Responding to C and D (see Fig. 2 and also Extended Data Fig. 2: second-order training) did not differ across SOC (NpHR: for day 1, $\chi^2(1,31)=0.6$, $P=0.439$; for day 2, $\chi^2(1,31)=0.29$, $P=0.593$; eYFP: for day 1, $\chi^2(1,21)=0.11$, $P=0.739$; for day 2, $\chi^2(1,21)=1.8$, $P=0.180$), there was no effect of trials (NpHR: for day 1, $\chi^2(2,47)=0.25$, $P=0.883$; for day 2, $\chi^2(2,47)=0.5$, $P=0.779$; eYFP: for day 1, $\chi^2(2,20)=0.45$, $P=0.798$; for day 2, $\chi^2(2,20)=0.36$, $P=0.834$), and no differences between C and D on each of the trial blocks (NpHR: for day 1, max. $\chi^2(1,31)=0.82$, $P=0.366$; for day 2, max. $\chi^2(1,31)=0.5$, $P=0.480$; eYFP: for day 1, max. $\chi^2(1,21)=0.1$, $P=0.739$; for day 2, max. $\chi^2(1,21)=2.0$, $P=0.157$). Similarly, responding to A after C or D did not differ (NpHR: for day 1, $\chi^2(1,31)=0.29$, $P=0.593$; for day 2, $\chi^2(1,31)=0.08$, $P=0.782$; eYFP: for day 1, $\chi^2(1,21)=0.11$, $P=0.739$; for day 2, $\chi^2(1,21)=0.11$, $P=0.739$), there was no effect of trials (NpHR: for day 1, $\chi^2(2,47)=0.57$, $P=0.752$; for day 2, $\chi^2(2,47)=0.37$, $P=0.832$; eYFP: for day 1, $\chi^2(2,32)=1.09$, $P=0.581$; for day 2, $\chi^2(2,32)=2.59$, $P=0.273$) or any differences on each of the trial blocks (NpHR: for day 1, max. $\chi^2(1,31)=2.57$, $P=0.109$, for day 2, max. $\chi^2(1,31)=0.4$, $P=0.527$; eYFP: for day 1, max. $\chi^2(1,11)=0.82$, $P=0.366$; for day 2, max. $\chi^2(1,21)=2.78$, $P=0.096$). Therefore, responding to A was combined.

Second-order probe test. After this training, rats received a probe test in which cues C and D were each presented six times in the absence of any reinforcement (see Fig. 2 and also Extended Data Fig. 2: probe test). Responding during the first trial of the second-order probe test was not normally distributed (NpHR: for C, $P<0.001$; for D, $P<0.001$; eYFP: for C, $P=0.002$; for D, $P<0.001$), so Wilcoxon's signed rank test was used to analyze differences between the conditions (see the legend to Fig. 2). Responding to C was lower compared with that to D in NpHR rats ($z=1.9$, $P=0.03$, $r=0.34$), but not in eYFP rats ($z=0.140$, $P=0.444$, $r=0.03$). In addition to analyzing the first trial of test, we also examined responding across the whole test, which confirmed the effects reported on trial 1. Responding across the entire test was generally not normally distributed (NpHR: for C, $P=0.001$, for D, $P=0.213$; eYFP: for C, $P=0.030$, for D, $P=0.003$). Wilcoxon's signed rank test confirmed that the difference between C and D persisted across the entire SOC test for NpHR rats ($z=2.38$, $P=0.009$, $r=0.421$) and the lack of difference for eYFP rats ($z=0.153$, $P=0.439$, $r=0.033$).

As mentioned earlier (see Main text), the differential effects of VTA DA shunting during A in SOC and blocking, as well as the lack of a difference in the eYFP rats in the second-order test, provide evidence that the disruptive effect of halorhodopsin on VTA DA signaling during second-order learning is not due to light artifacts serving to hinder processing of A. If it were, then we would see a disruption of the blocking effect as well (that is, learning about X). These results support temporal difference accounts by providing causal evidence that cue-evoked dopamine transients function as prediction errors.

Second-order replication experiment (Concordia). Rats received 20 daily conditioning trials between A, a visual cue (flashing light or steady light), and two sucrose pellets (2US) across 17 d. On days 18–20 the rats received 10 A→2US as well as 10 lever→2US conditioning trials. Pavlovian lever→2US conditioning was done to maintain high levels of responding during the subsequent phases of the study, that is, across SOC and test, six daily lever→2US trials were given interleaved with the critical X→A and Y→A SOC trials and the non-reinforced X and Y test trials. Responding to the lever is not of interest and was not therefore reported.

Conditioned responding for this experiment was reported using three measures: percentage time spent in the magazine during the cue (as described above), cumulative head entries during the cue period across a session and percentage trials with a head entry relative to all trials. Conditioned responding to A was normally distributed for the cumulative head entries measure (days 1–7 pre-tether: $P=0.473$; days 8–20 posttether: $P=0.104$) and percentage trials with a head entry (days 1–7 pre-tether: $P=0.842$; days 8–20 posttether: $P=0.112$) but not for percentage time spent in the magazine (days 1–7 pre-tether: $P=0.017$; days 8–20 posttether: $P=0.048$). A within-subject ANOVA revealed a linear trend across days (cumulative head entries: days 1–7 pre-tether, $F_{1,9}=63.78$, $P<0.001$; days 8–20 posttether, $F_{1,9}=7.66$, $P=0.022$; percentage trials with head entries: days 1–7 pre-tether, $F_{1,9}=34.26$, $P<0.001$; days 8–20 posttether, $F_{1,9}=7.52$, $P=0.023$). A Mann–Kendall test for percentage time spent in the magazine also reported an

increase in responding to A across days (days 1–7 pre-tether: $P<0.001$; days 8–20 posttether: $P<0.001$).

SOC took place during the subsequent 2 d and was identical to that described above with one exception: the lever continued to be paired with sucrose pellets for a total of six trials distributed across the SOC trials. Responding during this phase (see Extended Data Fig. 3) was normally distributed for all cues using percentage trials with a head entry (for C, $P=0.140$; for D, $P=0.198$; for A, $P=0.406$), but not for cumulative head entries (for C, $P=0.064$; for D, $P=0.044$; for A, $P=0.578$), or for percentage time spent in the magazine (for C, $P<0.001$; for D, $P<0.001$; for A, $P=0.244$). A Student's t-test revealed no differences in percentage trials with head entries between C and D (day 1: $t_9=0$, $P=1.0$; day 2: $t_9=0.198$, $P=0.847$). Wilcoxon's signed rank test was used to analyze the cumulative head entries for C and D during this phase (day 1: $z=0.526$, $P=0.599$; day 2: $z=0.281$, $P=0.779$). Friedman's ANOVA was used to analyze percentage time spent in the magazine during this phase for C and D. There was no difference between C and D across SOC (for day 1, $\chi^2(1,19)=0.11$, $P=0.739$; for day 2, $\chi^2(1,19)=0.14$, $P=0.706$), there was an effect of trials for day 1 ($\chi^2(2,29)=6.08$, $P=0.048$) but not for day 2 ($\chi^2(2,29)=2.77$, $P=0.250$), and no differences between C and D on each of the trial blocks (for day 1: max. $\chi^2(1,19)=0.14$, $P=0.706$; for day 2: max. $\chi^2(1,19)=0.67$, $P=0.414$). Similarly, responding to A after C or D did not differ across any of the measures on any of the days (cumulative head entries: day 1, $t_9=1.116$, $P=0.293$; day 2, $t_9=1.035$, $P=0.327$; percentage trials with a head entry: day 1, $t_9=0.208$, $P=0.840$; day 2, $t_9=0.176$, $P=0.864$). For percentage time spent in the magazine, responding to A after C or D did not differ (for day 1: $F_{1,9}<1$, $P=0.378$; for day 2: $F<1$, $P=0.919$), there was an effect of trials on day 1 ($F_{1,9}=9.20$, $P=0.014$) but not on day 2 ($F<1$, $P=0.714$) and no interactions (for day 1: $F<1$, $P=0.542$; for day 2: $F<1$, $P=0.704$).

After second-order training, rats received a probe test in which cues C and D were each presented six times in the absence of any reinforcement, but in the presence of six reinforced lever trials distributed across the C and D non-reinforced trials. Responding during the second-order probe test was generally not normally distributed (cumulative head entries: for C, $P<0.001$; for D, $P=0.106$; percentage trials with head entries: for C, $P<0.001$; for D, $P=0.067$; percentage time spent in magazine: for C, $P<0.001$; for D, $P=0.012$), so Wilcoxon's signed rank test was used to analyze differences between the conditions. Responding to C was lower compared with responding to D (cumulative head entries: $z=2.214$, $P=0.013$, $r=0.50$; percentage trials with head entry: $z=2.264$, $P=0.012$, $r=0.51$; percentage time in magazine: $z=2.197$, $P=0.014$, $r=0.49$).

Finally, we carried out additional modeling examining the effect of different strengths of inhibition (0 for no inhibition, 0.5 for partial inhibition, 1 for full inhibition) on learning to the cues in blocking (X, Y and Z) and SOC (C and D) in each of our models: the prediction model (Model V), the prediction-error model (Model Error) and a control for which all η values were 0. These data are captured in Extended Data Fig. 4.

Histology. The rats were euthanized with an overdose of carbon dioxide (NIDA) or sodium pentobarbital (Euthanyl) and perfused with phosphate-buffered saline followed by 4% paraformaldehyde (Santa Cruz Biotechnology Inc.). Fixed brains were cut in 40-µm sections, and images of these brain slices were acquired and examined under a fluorescence microscope (NIDA: Olympus Microscopy; Concordia: Carl Zeiss Microscopy). The viral spread and optical fiber placement (see Extended Data Figs. 2 and 3) were verified and later analyzed and graphed using Adobe Photoshop.

Data collection and statistics. Data that we collected using Colbourne Instruments or Med-Associates automated software and the text file output were analyzed using a custom-made script in MATLAB (Mathworks) or a custom-made Excel macro courtesy of S. Cabilio (Concordia University). Data from each phase of the experiments were checked for normality using the Shapiro–Wilk test in SPSS. In cases where the data were normally distributed, parametric tests were conducted. In cases where the data were not normally distributed non-parametric tests were used. As we tested specific hypotheses based on our modeling results, the directionality of the data were predetermined. Therefore, we used analysis of variance (ANOVA) and planned orthogonal contrasts in PSY2000 for parametric tests, and Friedman's ANOVA, Wilcoxon's signed rank test and the Mann–Kendall test for non-parametric analyses. Non-parametric effect sizes (r) were calculated for the probe tests following ref. [19]. Grubb's test was used to check for outliers.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Behavioral data will be made available upon reasonable request.

Code availability

Simulations were performed using custom-written functions in MATLAB (Mathworks), which are posted on Github (https://github.com/mphgardner/Basic_Pavlovian_TDRL/tree/Maes_2018).

References

14. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT Press, 1998).
15. Rescorla, R. A. & Wagner, A. R. in *Classical Conditioning: II. Current Research and Theory* (eds Black A. H. & Prokasy W. F.) 64–99 (Appleton–Century–Crofts, 1972).
16. Sharpe, M. J. & Killcross, A. S. The prelimbic cortex contributes to the down-regulation of attention toward redundant cues. *Cereb. Cortex* **24**, 1066–1074 (2014).
17. Mahmud, A., Petrov, P., Esber, G. R. & Iordanova, M. D. The serial blocking effect: a testbed for the neural mechanisms of temporal-difference learning. *Sci. Rep.* **9**, 5962 (2019).
18. Steinberg, E. E. et al. A causal link between prediction errors, dopamine neurons and learning. *Nat. Neurosci.* **16**, 966–973 (2013).
19. Olejnik, S. & Algina, J. Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychol. Methods* **8**, 434–447 (2003).

Acknowledgements

This work was supported by the Intramural Research Program at the NIDA; the Canada Research Chair's program (to M.D.I.); a Natural Sciences and Engineering Research Council of Canada Discovery Grant (to M.D.I.); a Natural Sciences and Engineering Research Council of Canada Undergraduate Student Research Award (to E.J.P.M.); and a Concordia University Undergraduate Research Award (to A.A.U.). The opinions expressed in this article are our own and do not reflect the view of the NIH/DHHS.

Author contributions

E.J.P.M., M.J.S., G.S. and M.D.I. conceived and designed the experiments. E.J.P.M., A.U. and M.L. carried out the surgical procedures and collected the behavioral data. C.Y.C., E.J.P.M., A.U. and L.M. supervised the immunohistologic verification of virus expression and fiber placement. M.P.H.G. conducted the computational modeling. M.J.S. and M.D.I. analyzed the data. G.S. and M.D.I. interpreted the data and wrote the manuscript with input from the other authors.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41593-019-0574-1.

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-019-0574-1.

Correspondence and requests for materials should be addressed to G.S. or M.D.I.

Peer review information *Nature Neuroscience* thanks S. Floresco and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

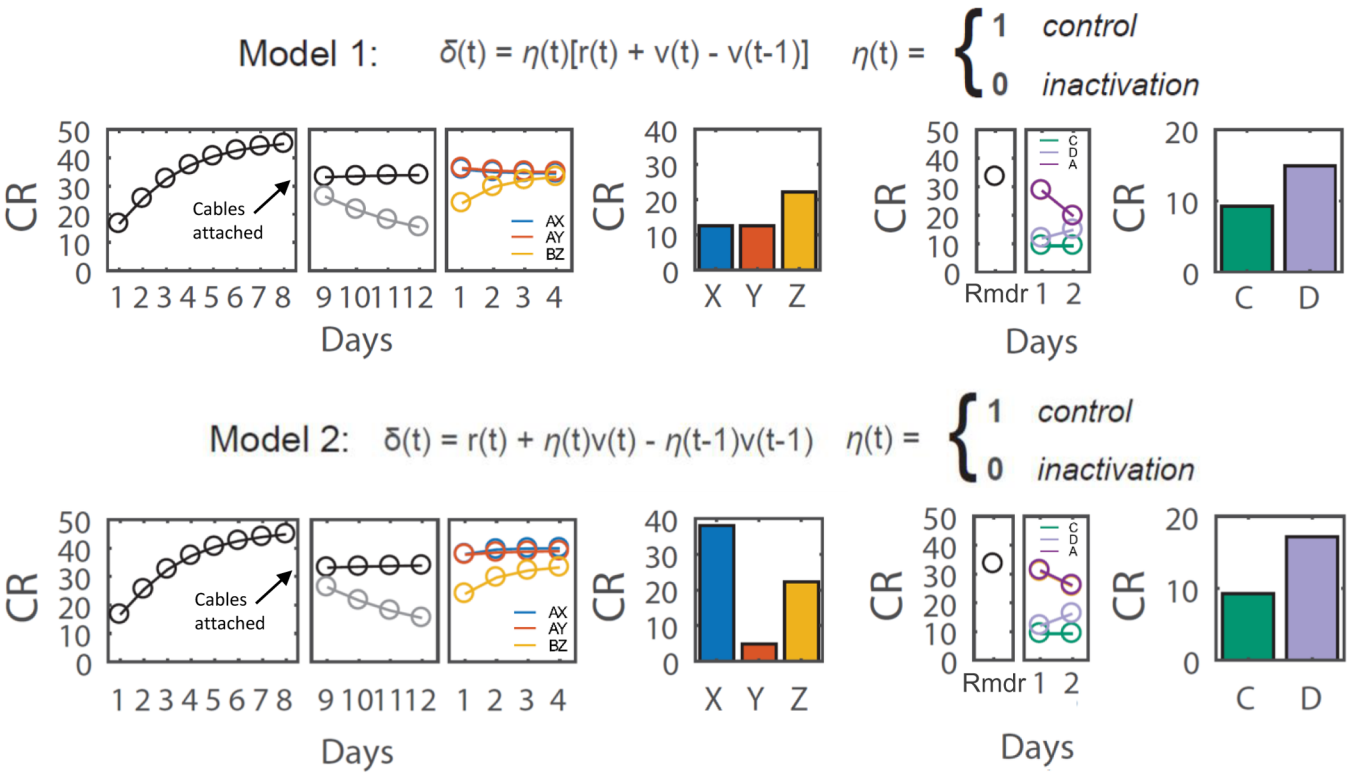
Reprints and permissions information is available at www.nature.com/reprints.

Blocking

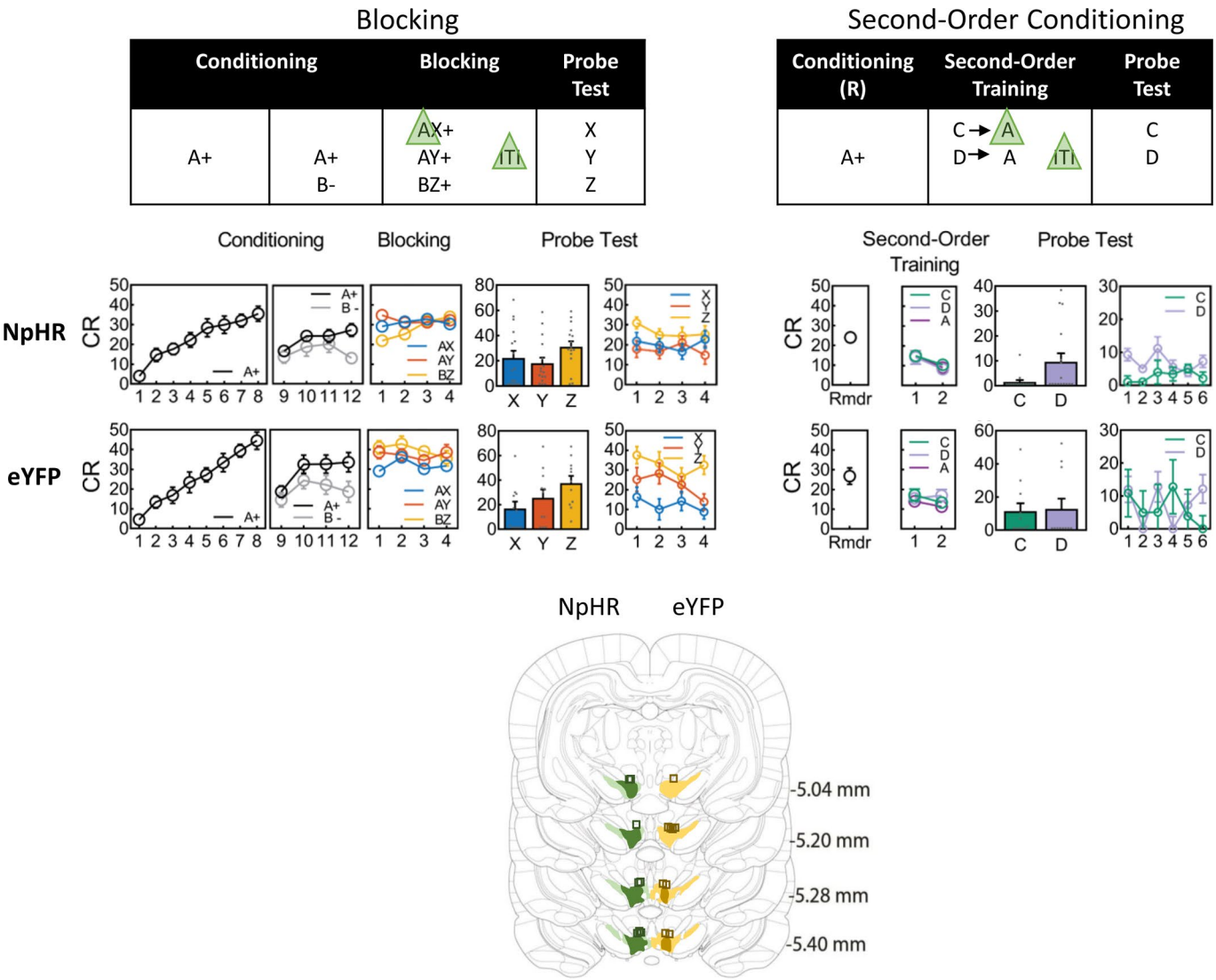
Conditioning		Blocking	Probe Test
A+	A+ B-	AX+ AY+ BZ+	X Y Z

Second-Order Conditioning

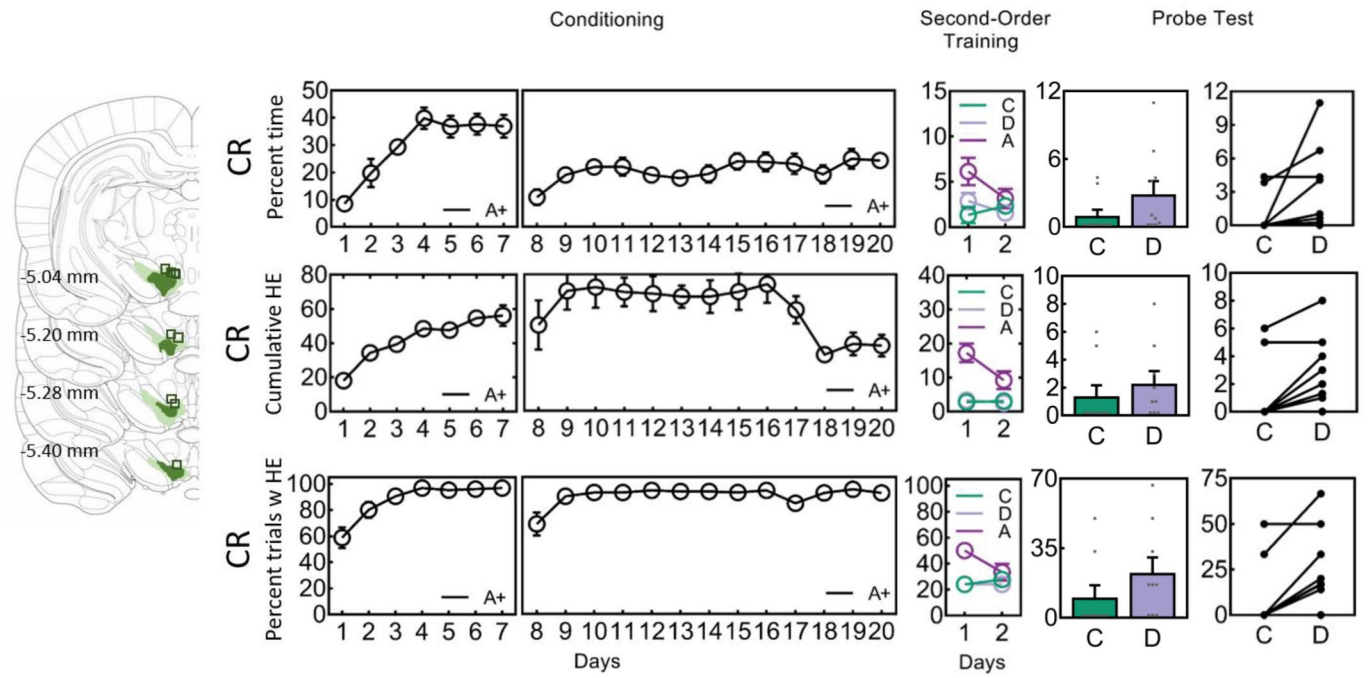
Conditioning (R)	Second-Order Training	Probe Test
A+	C→A D→A	C D



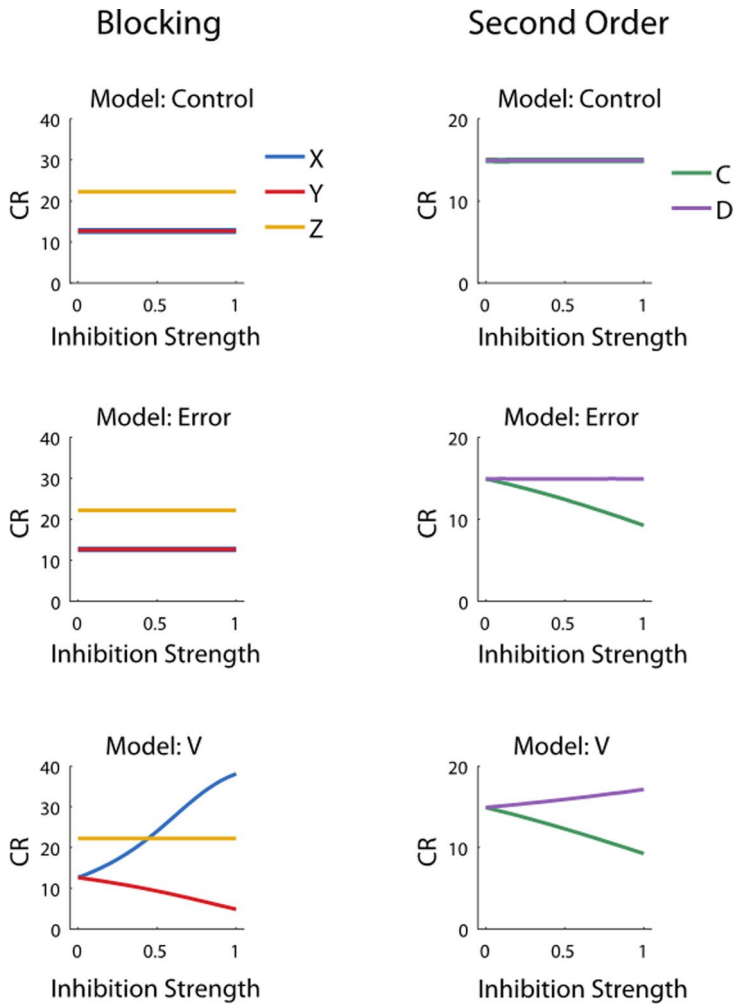
Extended Data Fig. 1 | Complete Designs and Modeling Data. Experimental design for within-subjects blocking and second-order conditioning as used in our study, along with graphs modeling the predicted results of shunting of the dopamine transient at the start of the reward-predictive cue, A, in each procedure. In Model 1 the VTA DA signal encodes a prediction error and in Model 2 it encodes a reward prediction. Bar graphs are reproduced from Fig. 1 in the main text; other panels model results of training in the other phases. Note the output of the classic TDRL model was converted from V to conditioned responding (CR) to better reflect the behavioral output actually measured in our experiments. The major impact of the neural manipulation was on responding to X in Model 2. Elimination of the prediction on AX trials in this model causes a positive prediction error on reward delivery in the blocking phase. This results in unblocking of X.



Extended Data Fig. 2 | ***Experimental designs for within-subjects blocking and second-order conditioning as used in our study with NpHR and eYFP rats.*** During Conditioning, responding to A but not B increased across days, and this responding was higher for A compared to B. During Blocking, responding to the control compound (DZ) was lower compared to blocking compound (AX, AY) at the start, but equivalent by the end of training, with no difference between the blocking compounds. Responding during the first trial of the Probe Test showed evidence of blocking (X and Y vs. Z) and no difference between the blocking cues (X vs Y, see Fig. 3 legend for statistics). Differences disappeared on subsequent trials. Responding to the retrained cue A increased across reminder (Rmdr) trials while that to C (i.e., C→A trials) and D (i.e., D→A trials) did not differ across second-order training. On Probe Test, responding to C was lower compared to D (see Fig. 2 legend for statistics) on the first trial as well as across the entire test. eYFP: The pattern of data obtained for the NpHR rats was similar to that obtained for the eYFP rats with one critical exception: there was no difference between C and D on Probe Test in the eYFP rats. Some data are reproduced from Figs. 2 and 3 in the main text. CR or conditioned responding is percent time spent in the magazine during the last 5 s of the cue. Drawings to the left illustrate the extent of expression of NpHR and eYFP and location of fiber tips within VTA.



Extended Data Fig. 3 | ***The cue-evoked dopamine transient is necessary for second-order conditioning in naïve rats.*** Drawings to the left illustrate the extent of expression of NpHR and location of fiber tips within VTA. The three panels of behavioral responding show behavioral data across the three phases of the second-order conditioning experiment represented using three different CRs (top—percent time spent in the magazine; middle—cumulative head entries during the CS across a single day of training; bottom—percent trials containing a head entry). Behavioral responding during A increased during Conditioning (see methods for statistics). Responding to C (i.e., C→A trials) and D (i.e., D→A trials) did not differ (see methods) during second-order training when shunting of VTA transients took place at the start of the reward-predictive cue, A. On Test, responding to C was lower compared to D (see methods for statistics for each of the CRs), showing that inhibition of the VTA DA signal at the start of A prevented A from supporting second-order conditioning to C whereas identical inhibition during the ITI left learning to D intact.



Extended Data Fig. 4 | ***Modeling data for the Blocking and Second-order experiments with different strengths of neuronal inhibition.*** The modeling data show how different inhibition strength (i.e., $\eta = 0$, 0.5, 1 as used in the models, see also Figure S1) affects the predicted conditioned responding on Probe Test across the different models. Model Control represents eYFP controls in which inhibition is not effective. Model Error represents the dopamine signal acting as a prediction-error in which increases in inhibition strength do not affect conditioned responding to X in blocking but lead to reduced conditioned respdoning to the C in second-order conditioning. Model V represents the dopamine signal as prediction in which increases in inhibition strength lead to greater conditioned responding to X in blocking (i.e., unblocking) and reduced conditioned responding to C in second-order conditioning.

nature research

Corresponding author(s): M.D. Iordanova

Last updated by author(s): Nov 10, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data was collected using equipment purchased from Coulbourn Instruments.
Data analysis	The software used to analyze data include IBM SPSS24, Matlab; Adobe Photoshop; open source: PSY2000, ai-therapy statistics tools; GraphPad Grubbs1 outlier calculator; Custom Excel Macros.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Behavioural data will be made available upon request. Simulations were performed using custom-written functions in MATLAB (Mathworks, Natick, MA), which are posted on Github (https://github.com/mphgardner/Basic_Pavlovian_TDRL/tree/Maes_2018).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Samples sizes were chosen based on published work, see references 6,8,9
Data exclusions	Four rats were excluded from the Concordia-bred cohort due to no virus expression (n=1), failure to consume the pellets during conditioning (n=1) or failure to receive stimulation due to broken ferrules (n=1) or cables (n=1). Four rats were excluded from the eYFP group due to no virus expression (n=2), failure to receive stimulation due to broken cables (n=1) and due to a significant outlier result accoding to Grubb's test.
Replication	Second-order experiment was replicated in naive rats. Data provided.
Randomization	We used a within-subjects design, which means that all subjects were tested in all conditions, so randomization was not required.
Blinding	Data were collected using an automated system and all subjects were tested in all conditions within the same automated session (within subjects design), therefore blinding does not apply.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Th-cre rats on Long Evans background, males, females, 3months of age
Wild animals	This study did not involve wild animals.
Field-collected samples	This study did not involve data collected form the field.
Ethics oversight	All experimental procedures conducted at the NIDA-IRP were in accordance with the Institutional Animal Care and Use Committee of the US National Institute of Health guidelines, and those conducted at Concordia University were in accordance were treated in accordance with the approval granted by the Canadian Council on Animal Care and the Concordia University Animal Care Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.