Behavioral/Systems/Cognitive

Tripartite Mechanism of Extinction Suggested by Dopamine Neuron Activity and Temporal Difference Model

Wei-Xing Pan,[1] Robert Schmidt,[2] Jeffery R. Wickens,[2,3] and Brian I. Hyland[1]

Departments of [1]Physiology and [2]Anatomy and Structural Biology, School of Medical Sciences, University of Otago, Dunedin 9054, New Zealand, and [3]Okinawa Institute of Science and Technology, Okinawa 904-2234, Japan

Extinction of behavior enables adaptation to a changing world and is crucial for recovery from disorders such as phobias and drug addiction. However, the brain mechanisms underlying behavioral extinction remain poorly understood. Midbrain dopamine (DA) neurons appear to play a central role in most acquisition processes of appetitive conditioning. Here, we show that the responses of putative DA neurons to conditioned reward predicting cues also dynamically encode two classical features of extinction: decrement in amplitude of previously learned excitatory responses and rebound of responding on subsequent retesting (spontaneous recovery). Crucially, this encoding involves development of inhibitory responses in the DA neurons, reflecting new, extinction-specific learning in the brain. We explored the implications of this finding by adding such inhibitory inputs to a standard temporal difference model of DA cell activity. We found that combining extinction-triggered plasticity of these inputs with a time-dependent spontaneous decay of weights, equivalent to a forgetting process as described in classical behavioral extinction literature, enabled the model to simulate several classical features of extinction. A key requirement to achieving spontaneous recovery was differential rates of spontaneous decay for weights representing original conditioning and for subsequent extinction learning. A testable prediction of the model is thus that differential decay properties exist within the wider circuits regulating DA cell activity. These findings are consistent with the hypothesis that extinction processes at both cellular and behavioral levels involve a dynamic interaction between new (inhibitory) learning, forgetting, and unlearning.

Key words: **dopaminergic neuron; extinction; classical conditioning; extracellular recording; temporal difference model; reward learning**

Introduction

Extinction of behavioral responses is essential for adaptive flexibility in the face of changing environmental contingencies. Animals learn to respond to arbitrary stimuli if these are associated with reward or punishment. If the reward or punishment is then discontinued the behavioral responses to the stimulus undergo a characteristic sequence of changes leading to a decline in responding, a process known as extinction. It is generally believed that extinction represents new, inhibitory learning, rather than unlearning or forgetting (Pavlov, 1927; Konorski, 1948). This hypothesis is supported by several behavioral phenomena, including spontaneous recovery, rapid reacquisition, renewal and reinstatement (Bouton and Moody, 2004). At the neural level, previous studies showed that conditioned excitatory responses in several brain areas can be reduced or abolished with extinction (Diamond and Weinberger, 1986; Komura et al., 2001; Toyomitsu et al., 2002). However, such reductions would also be consistent with a simple "undoing" of conditioning. Direct evidence of active inhibitory processes assumed to underlie extinction is currently lacking.

Recent studies have shown that the patterns of midbrain dopamine (DA) neuron activities share several key features of conditioning, such as acquisition (Schultz, 1998; Pan et al., 2005), generalization (Mirenowicz and Schultz, 1996), blocking phenomena (Waelti et al., 2001), and conditioned inhibition (Tobler et al., 2003). The features of DA activity during this learning match the prediction-error signal of the temporal difference (TD) algorithm for machine learning, suggesting that DA neurons might provide such a prediction-error signal in the brain (Montague et al., 1996; Schultz et al., 1997; Pan et al., 2005). Given that DA neural activity strongly links to processes of reward-mediated associative learning, it has been postulated that DA neurons may also be involved in extinction learning (Myers and Davis, 2002). However, this has not been directly tested. Furthermore, existing TD models predict the eventual loss of conditioned DA neuronal responses during extinction (Montague et al., 1996), but do not contain any intrinsic mechanism that could produce spontaneous recovery or any other specific phenomenon of extinction.

In the present study, we characterized DA cell activity during extinction, and used the findings obtained to develop an ex-

Received Jan. 18, 2008; revised Aug. 14, 2008; accepted Aug. 18, 2008.

This work was supported by the Neurological Foundation of New Zealand, Lottery Grants Board, and Marsden Fund. Thanks to Prof. Gordon Arbuthnott for comments on this manuscript and Prof. Neil McNaughton for statistical advice.

Correspondence should be addressed to Brian I. Hyland, Department of Physiology, School of Medical Sciences, University of Otago, Dunedin 9054, New Zealand. E-mail: brian.hyland@otago.ac.nz.

W.-X. Pan's present address: National Institutes of Health–National Institute on Drug Abuse, Behavioral Neuroscience Branch, 5500 Nathan Shock Drive, Baltimore, MD 21224.

R. Schmidt's present address: Bernstein Center for Computational Neuroscience, Institute for Theoretical Biology, Humboldt University, D-10099 Berlin, Germany.

DOI:10.1523/JNEUROSCI.0255-08.2008

tended TD model that successfully simulated the DA cell activities during associative learning and extinction. The requirements of the model are consistent with mechanisms for extinction in the brain which incorporate three distinct processes: Pavlov's original conception of new inhibitory learning, a passive forgetting process, as well as an underlying unlearning of previous conditioning.

Materials and Methods

Data set. The results described here represent the outcome of analysis of a data set of DA neurons, including 16 for which the responses to conditioned cue signals during acquisition of cue–reward associations have been reported previously (Pan et al., 2005), and another two subsequently recorded cells that have not been described previously. For entry to this analysis, all cells had to meet our electrophysiological and pharmacological criteria for dopamine cells, and also had to have been tested with extinction procedures. The presumed DA cells reported here were obtained from 13 rats of a group of 58. For comparison, we included another 28 cells, recorded from the wider group of rats, which either were considered nondopaminergic or in which the classification was ambiguous.

Unit recordings. All procedures were approved by the University of Otago Animal Ethics Committee. Male Wistar rats weighing 250–400 g were used. Under full anesthesia (sodium pentobarbital, 60 mg/kg, i.p.), a bundle of eight microwire recording electrodes (0.001 inch Formvar-insulated nichrome; A-M Systems) was implanted in the dopamine cell groups of the substantia-nigra pars compacta or the ventral tegmental area (VTA; 5.0–5.5 mm posterior to bregma, 0.5–2.0 mm lateral to midline, and 6.5–7.0 mm below the surface of skull) (Paxinos and Watson, 1997). The microelectrodes were mounted on an on-head microdrive, which allowed advance of the electrodes during the subsequent recording sessions.

For neural recordings, signals from the electrodes were amplified (2000–10,000×) and filtered (0.2–10 kHz bandpass) using a custom-built multichannel signal conditioner. Channels with cellular activity were digitized (20 kHz) and recorded on computer using Discovery software (Datawave Technologies). The extracellularly recorded action potentials were discriminated from each other and from noise based on wave shape using the spike sorting features of Datawave Personal Scientific Workstation software. Putative dopamine cells were identified using the previously published criteria for cells recorded with these techniques (Hyland et al., 2002; Pan and Hyland, 2005; Pan et al., 2005); the initial selection was determined by low firing rate (<10 Hz) and relatively broad action potential (>1.0 ms). However, only cells that were also profoundly ($>50\%$) inhibited by the dopamine agonist apomorphine (750 μg/kg, i.p.) or the D_2-receptor selective agonist quinpirole (400 μg/kg, s.c.) were accepted into the group of presumed dopamine cells (Aebischer and Schultz, 1984). Dopamine agonist tests were performed after initial conditioning experiments; additional behavioral testing was not performed until at least 2 h after behavioral effects had ceased and cell activity returned to the baseline firing rate.

Histology. To confirm location of cells, after completion of the experiments lesions were produced at the tip of a wire from which recordings had been made by passing direct current (9 V for 1–2 min). After 5–10 d survival time, rats were killed by anesthetic overdose, perfused with saline and then formalin solution, the brains sectioned on a freezing microtome, and the position of the marking lesions and cannula tracks mapped on standard atlas sections (Paxinos and Watson, 1997).

Behavioral paradigms. After a 1 week recovery from surgery, rats were fluid deprived for 24 h and then brought to the recording chamber, a Perspex box with a floor area of 25 × 16.5 cm located in a quiet, darkened room. Fluid deprivation was maintained by restricting *ad libitum* access to water in the home cage to 5–10 min per day. Rats were weighed daily, and restriction eased if they fell to $<85\%$ of their projected weight–growth curve. The recording chamber included a fluid spout, Med Associates Sonalert speaker, and a house light. Licks were detected when the tongue broke an infrared beam across the spout opening. Fluid was delivered to the spout by transient activation of a solenoid, which generated an audible click. Rats were familiarized with the recording chamber and the association of solenoid click with delivery of fluid before making cell recordings.

Automated behavioral paradigms were controlled via Med-PC interfaces using programs written in MedState notation (Med Associates). Cells were tested with conditioning, prediction error, and extinction paradigms. For the majority of cells included in this study, the effect of conditioning and prediction error testing has been described previously (Pan et al., 2005). The conditioning paradigms consisted of an "unsignaled reward" and a "signaled reward." For the unsignaled reward, small volumes (≈0.05 ml) of water sweetened with saccharin (0.005 M solution) were delivered to the spout by activating the solenoid valve at pseudorandom delays (10–20 s). The random delay period began once the previous reward had been retrieved, indicated by licks detected at the spout. Rats learned this task before recordings began, so the solenoid click already represented a conditioned signal. This paradigm was routinely run each day while checking for the presence of cells, in case it activated otherwise silent or extremely slow firing cells. Signaled reward paradigms conditioned an association between tone cues and solenoid activation for reward delivery. These were generally only run when suspected DA cells were recorded. Here, solenoid activation was preceded by one or two 4.5 kHz tone cues of 0.5 or 2 s duration. Two-cue tests involved either 2 or 0.5 s cues, with an intercue interval of the same duration as the cue, and reward delivered at the end of the second cue. For one-cue tests, the cue was always 0.5 s in duration with the reward either delivered at the end of the cue (delay conditioning) or 1 s after (trace conditioning). Importantly, regardless of the design used, most DA cells developed a short latency response to cue onset (Pan et al., 2005).

For the main aim of the present study, we examined the responses of DA cells during extinction of the previously conditioned associations using two paradigms, "cues only" and "solenoid only." In the cues-only paradigm the same cue tone as used in the signaled reward paradigm was delivered from the speaker at pseudorandom intervals, but with no solenoid activation. Trials proceeded with pseudorandom intertrial intervals (10–20 s) with no requirement that the rat lick at the spout. Extinction sessions consisted of 50–100 trials. In a few cases of cells with particularly stable recordings, multiple extinction sessions were able to be recorded, including some sessions recorded the next day. A subset of cells was also tested for extinction of solenoid responses using the solenoid-only paradigm. For this, the same procedure as for unsignaled reward was run, but the fluid supply line was disconnected from the spout, and the released fluid drained to a beaker beside the task chamber. This ensured that the pitch and volume of the solenoid click was identical to the unsignaled reward condition.

Analysis of neuronal activity. Cell firing data were analyzed with peristimulus time histograms (PSTHs) centered on reward or cue delivery to show average activity over trials, and dot-raster displays to show individual trials. Baseline average firing rates were calculated for each cell from PSTH epochs. Cells were classified as showing excitatory responses to conditioned stimuli if two or more sequential 25 ms poststimulus bins were >2 SD above the baseline mean. Because of the floor effect for the amplitude of inhibitory troughs imposed by the low baseline firing rate of DA neurons, inhibitory troughs were assessed using a different criterion; cells were classed as showing a primarily inhibitory response if there was a $>10\%$ fall in firing rate for at least three successive bins. For quantitative analysis of extinction-evoked changes in responses among the excited cells, the postcue period was divided into two epochs, lasting from 0 to 125 ms and 125 to 250 ms after cue onset for tone cues, and 50 to 175 and 175 to 300 ms after solenoid click to allow for longer latency of solenoid responses. The longer latency for solenoid responses presumably reflected the different pitch, amplitude, and rise time of this stimulus, biophysical properties that can dramatically alter response latencies of neurons in the auditory pathway (Heil, 1997). For each epoch, change in firing rate after the cue was expressed as a modulation index e/b, where e is the mean firing rate in the epoch and b is the mean firing rate in the precue period. For cells tested with more than one sequential block of extinction training, we took the last block, and for cells that showed spontaneous recovery of responding (see below), analysis was limited to the last part of each data file.

Changes in population mean epoch values were analyzed using nonparametric statistics [Friedman's test with *post hoc* tests for planned comparisons (Dunn's multiple-comparison test) or Wilcoxon signed rank test] or one-sample *t* tests in GraphPad Prism version 4.00 for Macintosh.

Behavioral extinction. During the initial cell recordings, detailed data about lick events was not collected during cue-only extinction sessions; furthermore, different cells were tested with varying durations of extinction because of differences in the length of time for which they could be recorded. Therefore, to characterize the typical time course of behavioral extinction using these paradigms, a separate group of unoperated animals were first trained as above and then tested in three blocks of extinction. Behavioral conditioning was quantified per 25 trials within each block as the number of trials in which a lick response occurred between the onset of the cue and the activation of the solenoid. These data were statistically analyzed using ANOVA with extraction of orthogonal linear, quadratic, and cubic polynomial components of performance across trial blocks (SPSS).

To assess whether intact inhibitory transmission at the level of the dopamine cell fields is required for normal expression of behavioral extinction, an additional group of animals was tested in a single session of extinction after intracerebral injection of the $\mathrm{GABA_A}$ receptor antagonist bicuculline. Rats ($n = 12$) were implanted with bilateral 30-gauge stainless-steel guide cannulas aimed at the midbrain DA cells in the VTA (anteroposterior, 5.3 mm; lateral, 1.0 mm; depth, 6.5 mm) (Paxinos and Watson, 1997) and, after 5 d postoperative recovery, trained as described above to associate solenoid clicks with fluid reward over 3 d with 200 trials/d. Solenoid extinction tests (50 trials with solenoid activation but no fluid delivery) were performed from the fourth day. Extinction trials began 5 min after bilateral simultaneous infusion of either vehicle (saline; six rats) or bicuculline methochloride (Sigma-Aldrich; six rats). Injections were made by needles inserted through the guide cannulas so that the tips had a final stereotaxic depth of 8.5 mm. Infusion volumes were 0.5 μl on each side at 0.25 μl/min, controlled by two syringe pumps. The final bicuculline dose was individually adjusted for each rat to a level that produced only slight, if any, stereotypic behavior and ranged from 0.01 to 0.025 μg/0.5 μl. Responding during extinction was quantified (conditioned response index) as the number of trials per 10-trial block in which a lick response occurred within 10 s of the conditioned stimulus. A conditioned response index was calculated for each group of 10 trials as the percentage of trials with lick responses, which was arcsine transformed ($Y' = \text{arcsine}\ [\text{sqrt}(Y)]$) to normalize the variance before comparison of control and treated groups using ANOVA with extraction of orthogonal linear, quadratic, and cubic polynomial components of performance across trial blocks.

TD model. We made two additions to the basic TD algorithm that we previously applied to DA neural activity during learning (Pan et al., 2005), which in turn built on previous work (Sutton, 1988; Montague et al., 1996; Schultz et al., 1997). First, we implemented two sets of weights, one positive, that ranged from zero to increasingly positive values, and the other negative, ranging from zero to increasingly negative values (Klopf, 1988). We refer to the positive weights as "excitatory" and the negative weights as "inhibitory." Second, in addition to active weight changes driven by prediction error (δ), excitatory and inhibitory weights were also subject to spontaneous decay at each time step. We refer to this passive decay process as forgetting, to distinguish it from the active process inherent in standard TD whereby weights are reduced by negative prediction error. During TD, increases in weights driven by positive prediction error represent the learning of associations. We therefore refer to the reduction of weights by active TD processes as unlearning, i.e., the active erasure of learning. Both learning and unlearning rates are determined by one parameter, whereas another determines forgetting. Excitatory and inhibitory weights had separately defined parameter sets: α and β for learning and unlearning at excitatory and inhibitory weights, respectively, and ψ^+ and ψ^- for forgetting. We found that with appropriate adjustment of learning and forgetting rate parameters, the model achieved both extinction and learning phenomena (see below and Results).

In brief, the two sets of weights (see Fig. 6*a*) meant that the state vector x representing the stimulus was associated with both an excitatory and an inhibitory weight vector ($w^+(t)$ and $w^-(t)$, respectively). Estimates (predictions) of future rewards were formed by the dot product of the state and weight vectors separately for positive and negative weights, as $P^+(t) = x(t) \cdot w^+(t)$ and $P^-(t) = x(t) \cdot w^-(t)$, respectively, and the total reward prediction [$P(t)$] was formed by the sum of these predictions. The expected reward at each time step t was then calculated in the usual way as the temporal difference TD(t) of predictions at two successive time steps (Sutton, 1988). This expected reward was compared with the scalar reinforcement signal $r(t)$ to generate the prediction-error signal $\delta(t)$.

This prediction error was then used as a teaching signal to update the vector weights according to the weight change rules $\Delta w^+(t) = \alpha\delta(t)e(t)$ and $\Delta w^-(t) = \beta\delta(t)e(t)$, where α and β are the learning (and thus also unlearning) rates for excitatory and inhibitory weights, respectively, and e is the eligibility trace parameter. The eligibility trace $e(t)$ (Sutton, 1988) was calculated recursively by $e(t + 1) = \lambda e(t) + x(t)$. The eligibility trace decay parameter (λ) was set to 0.9, which we previously found to be optimal for enabling the model to replicate the pattern of DA neuron activity responses to cues and rewards during acquisition of conditioning (Pan et al., 2005).

Finally, spontaneous decay of excitatory and inhibitory weights (forgetting) was achieved by multiplying each weight by decay factors ψ^+ and ψ^- with $0 < \psi \leq 1$ at each time step. Thus, numerically smaller decay factors determined a faster rate of decay.

Results

Putative dopamine neurons were preselected on the basis of slow firing rate (less than $\approx$10 Hz) and relatively broad action potential (more than $\approx$1.2 ms total), and in all cases were tested with systemic injection of either the mixed agonist apomorphine or the $\mathrm{D_2}$-selective agonist quinpirole. Only those profoundly ($\geq$50%) inhibited in firing rate, over the same time course as the stereotypy induced by successful drug injection, were accepted as likely dopaminergic cells for the purposes of this study, and for the most part only such cells were tested with the full range of behavioral procedures. Overall, from $>$200 neurons encountered as the electrodes advanced, 28 met our criteria for being dopaminergic, and 18 of them (from 13 rats) were recorded during both signaled reward conditioning and cues-only extinction paradigms, and so were candidates for analysis of extinction phenomena. Figure 1 shows an analysis of the waveform amplitude and duration characteristics of all these 18 presumed DA cells, using measures recently described by Roesch et al. (2007), along with individual example waveforms and the dopamine agonist test histograms. The waveform amplitude ratio index was calculated as $n - p = n + p$, where n is the amplitude from baseline to first negative peak after a positive peak, and p is the amplitude from baseline to first positive peak. Waveform duration was characterized by the late component, taken from the first negative peak to the subsequent positive peak (or to baseline if no late positive peak was present). Generally, DA cell waveforms were characterized by an initial positive component and relatively broad trough. In contrast, a sample of other cells recorded from the same overall pool of rats tested with but not inhibited by apomorphine were generally characterized by lesser or absent initial positivity, as reported previously (Roesch et al., 2007). However, testing all cells with dopamine agonists revealed some overlap in the distributions of waveform parameters in our data set, with a small number of cells with large positive components and broad troughs rejected from the presumed DA set because of excitatory responses to agonists, and some presumed DA cells having little initial positivity. We therefore used response to dopamine agonists as the final arbiter of categorization. Such an approach might miss subpopulations of cortically projecting DA cells,

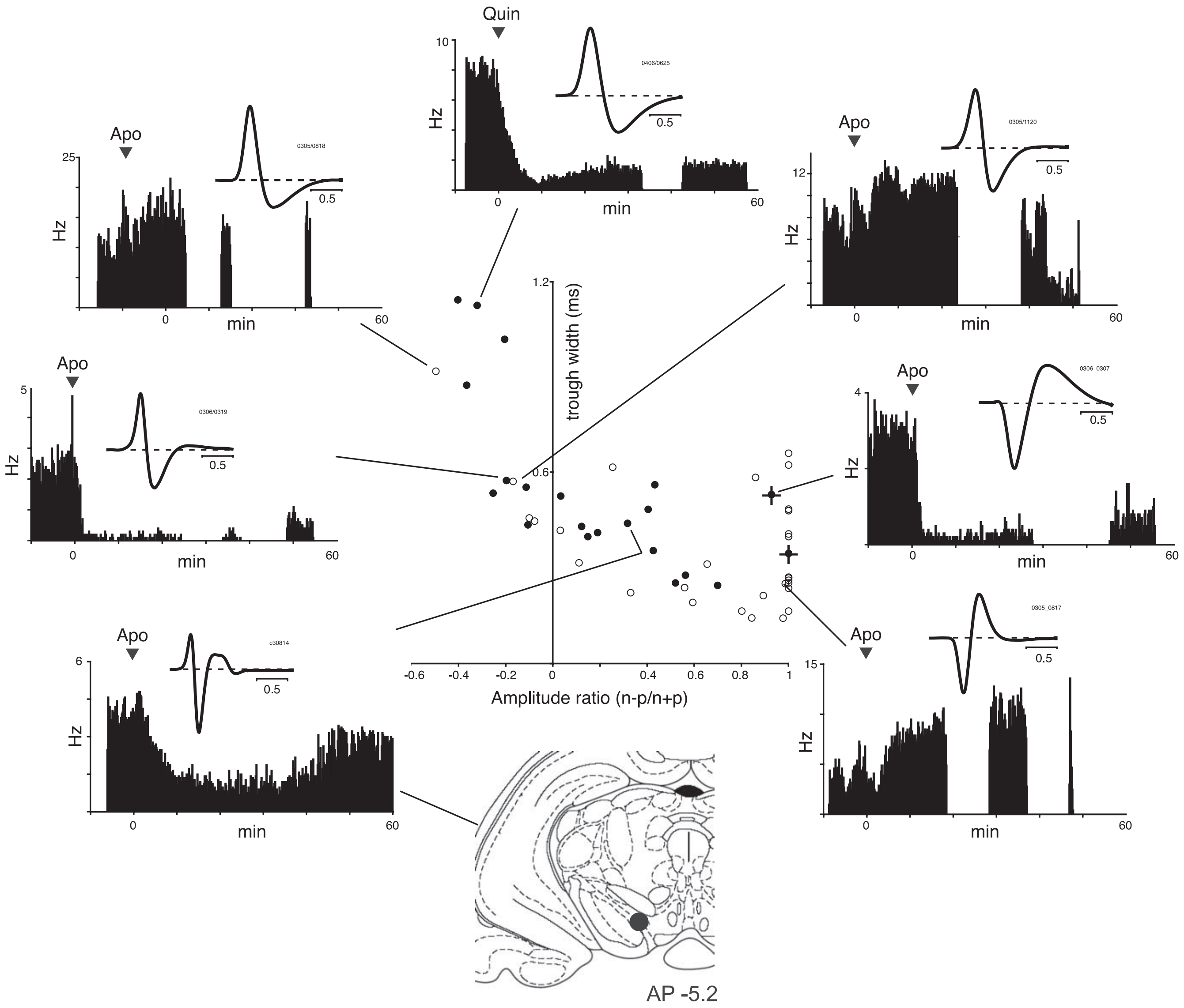

Figure 1. Characteristics of recorded neurons. Scatter plot shows results of an analysis of waveform shape using measures described by Roesch et al. (2007). Filled circles show presumed dopamine cells, on the basis of inhibitory response to dopamine agonist and slow firing rate. Circles with superimposed crosses indicate presumed dopamine cells that did not show trial-to-trial prediction error signaling (see supplemental material, available at www.jneurosci.org). Open circles show results for a sample of neurons that were not accepted as dopaminergic for the purposes of this study on the basis of either no response or an excitatory response to dopamine agonist. Horizontal axis, Waveform amplitude ratio index; vertical axis, waveform trough duration. Histograms show the firing rates of example cells from across the distribution before and after systemic injection of apomorphine (Apo) or quinpirole (Quin), with inset average waveforms. The atlas section (Paxinos and Watson, 1997) shows the recovered position of the indicated cell.

which do not strongly express D_2 receptors (Chiodo et al., 1984; Lammel et al., 2008), but has the advantage of being less likely to include non-DA cells in the sample.

Of the 18 DA cells, 15 showed phasic, short latency excitatory responses to conditioned cues that predicted reward, and were analyzed quantitatively for extinction effects. The remaining three cells included one that was not responsive to task stimuli, and two that showed inhibitory responses. Additional details of the characteristics of the presumed DA and non-DA cells included in Figure 1 are provided in the supplemental material (available at www.jneurosci.org), including response to overtrained cues, prediction-error signaling (supplemental Figs. S1, S3), lick-related activity of presumed non-DA cells (supplemental Figs. S2, S3), and rhythmic patterns of firing (supplemental Fig. S4, available at www.jneurosci.org as supplemental material).

Extinction training extinguishes conditioned responses of DA cells

All 15 cue–responsive DA cells were tested for cue extinction (cues-only paradigm), and five were additionally tested for extinction of responses to the solenoid click (solenoid-only paradigm). As illustrated by the examples in Figure 2, extinction training led to decreases in previously robust responses to these stimuli. Additional examples are also seen in Figures 4*a* and 5*a*.

The population analysis in Figure 3 shows that within the number of extinction training trials available (which differed from cell to cell), the majority of recorded cells exhibited such extinction effects. For this analysis, we broke the poststimulus time into two epochs to separately capture changes occurring in early and late components of the response (Fig. 3*a*, inset). Excitatory peaks occurring in signaled reward and unsignaled reward conditions were mostly encompassed within the first epoch (0–

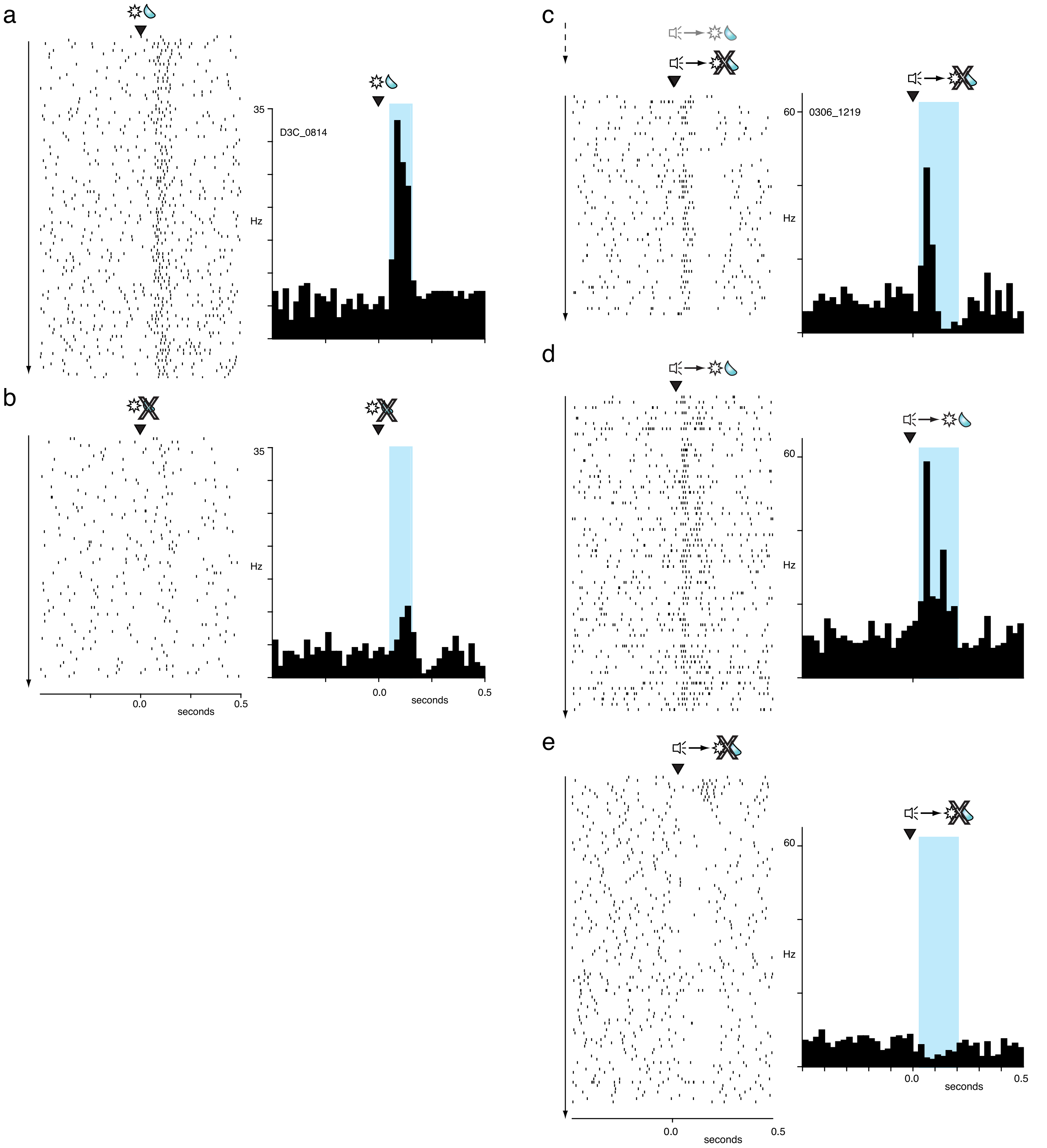

Figure 2. Loss of conditioned excitation and development of new inhibition in DA neurons during extinction training. ***a***, ***b***, Solenoid extinction. ***a***, Robust response of cell to sound of solenoid click associated with reward delivery in the unsignaled reward paradigm. The dot raster shows trials in time order (first trial in session at top). The star and droplet symbols indicate the time of solenoid activation delivering fluid reward. The histogram shows the averaged activity across all trials. Blue shading demarcates the duration of the excitatory peak. ***b***, Response of same cell as in ***a*** during extinction in a subsequent session of the solenoid only paradigm. Solenoid click (star symbol) was now not associated with fluid delivery. Blue shading shows the duration of the preextinction excitatory peak, from the histogram in ***a***. ***c–e***, Cue extinction. ***c***, Dashed arrow and gray speaker/star/droplet symbol indicate previous exposure of the animal to conditioning with the signaled reward paradigm the previous day, before this cell was encountered. The raster and histogram show the first recording of this cell, which was during the cues-only paradigm, i.e., an extinction session as far as the animal was concerned. Blue shading indicates duration of excitatory peak during conditioning (from the histogram in ***d***). ***d***, The same cell as in ***c***, showing retraining with the signaled reward on a subsequent session. ***e***, Extinction with cues only, after the retraining in ***d***.

125 or 50–175 ms poststimulus onset, respectively), as indicated by the reduced amplitude of the modulation index within the second epoch in these paradigms (Fig. 3*a*,*b*, conditioned).

Considering cue responses (Fig. 3*a*,*b*), the mean response amplitudes across all epochs and conditions varied significantly ($p < 0.0001$, Friedman's test), and planned comparisons showed that mean firing rate modulation in each epoch of the extinction training sessions was lower than the equivalent epoch in the conditioning sessions ($p < 0.01$, *post hoc* Dunn's multiple-comparison tests). Only one cell failed to show a clear diminution in responses. Interestingly, this cell was also atypical in another respect as it was one of two DA cells that did not respond to solenoid click in the unsignaled reward condition, similar to "nonreward" cells described recently by Roesch et al. (2007) (supplemental material, available at www.jneurosci.org). For cells tested with solenoid extinction, there was also a clear trend for reduction in response amplitudes (Fig. 3*c*,*d*) ($p < 0.001$ across all epochs and conditions, Friedman's test), but with the number of cells available the *post hoc* pairwise tests did not reach significance.

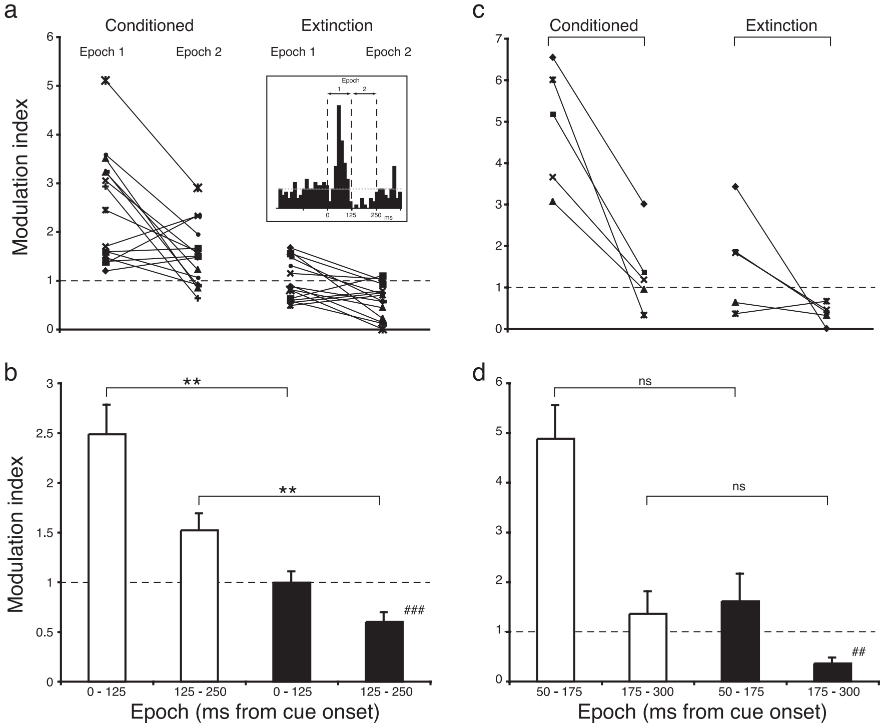

Figure 3. Quantitative analysis of DA neuron responses during extinction. ***a***, ***b***, Response to cue tones. ***a***, Graph shows data for all cells ($n = 15$) during conditioning, when cues predicted rewards, and extinction, when cues were not associated with reward. As indicated on the inset example PSTH, for this analysis the immediate postcue period was divided into two periods, epoch 1 from 0 to 125 ms and epoch 2 from 125 to 250 ms after cue onset (time 0 on the inset PSTH). For each epoch, change in firing rate from baseline (horizontal dotted line in the inset example PSTH) after the cue is expressed as a modulation index (see Materials and Methods); values >1 suggest excitation and <1 suggest inhibition of firing compared with baseline. In the main plot, lines connect data from the two epochs from each cell. The same cells occur in both conditioned and extinction columns, but these are not connected across columns for clarity. ***b***, Histogram shows mean + SEM of the data in ***a***, with conditioning epochs in white and extinction epochs in black. $^{**}p < 0.01$, *post hoc* Dunn's multiple-comparison test; $^{###}p < 0.001$, one-sample *t* test comparing mean to theoretical mean = 1. ***c***, ***d***, Response to solenoid. Graphs show data for responses to solenoid clicks (not preceded by cues) that either delivered (conditioned) or did not deliver (extinction) fluid reward. Details are as for ***a*** and ***b*** except that epochs were 50–175 and 175–300 ms after click to allow for longer latency of solenoid responses. ns, Not significant, *post hoc* Dunn's multiple-comparison test; $^{##}p < 0.01$, one-sample *t* test comparing mean to theoretical mean = 1.

DA neurons develop inhibitory responses during extinction learning

In several cells, such as those shown in Figure 2, we noted that extinction training was not only associated with reduction in amplitude of excitatory responses to cues, but also with development of an inhibitory component in the response profile. Quantitative analysis at the population level (Fig. 3) of all 15 cells tested for cue extinction showed that 12 (80%) had a reduction in mean firing rate of ≥10% below baseline in one or another epoch of the postcue period (Fig. 3*a*), and this was true for all five cells tested with solenoid extinction (Fig. 3*c*). The epoch in which this criterion was reached differed across the cells; in 4 of 15 cue extinction tests and 3 of 5 solenoid extinction tests, it was not until the second epoch. This difference at the single-cell level was reflected in the population means, with only the means for the second epoch lying below their respective baseline (Fig. 3*b*,*d*). One-sample *t* tests for the each of the second epochs of cue and solenoid extinction confirmed that in each case the mean normalized values were significantly below baseline ($p < 0.001$ and $p < 0.01$, respectively). The extension of the inhibitory response into the second epoch would be consistent with a longer duration, and/or later onset of an inhibitory input to the cells, compared with the excitatory input giving rise to the peaks.

Extinguished neuronal responses show spontaneous recovery

Spontaneous recovery is an important feature of behavioral extinction, which marks it as more than simply erasure of learning. To assess whether dopamine cell responses show this phenomenon, we took advantage of the fact that some cells were recorded in multiple blocks of extinction training (cues-only paradigm). In particular, we examined in detail a group of six cells that were recorded in two sessions on one day, and a third session on the next day, with no intervening signaled reward retraining. Each session of ≈100 trials was broken into sequential quarters, yielding 12 subblocks of ≈25 trials each. An example analysis for one cell is shown in Figure 4*a*. In this cell, a robust response to the cue onset persisted through much of the first block, but declined by the end. A peak is evident at the beginning of the next session, which declines more rapidly, and this sequence was repeated in the third session, by the end of which an inhibitory pause developed.

To quantify spontaneous recovery across the six cells, we calculated a "recovery index." For each subblock, the cue-induced modulation was quantified as the average activity in the 250 ms after the cue divided by the precue baseline. All data were then normalized to the modulation in the first subblock of extinction session 1. We then compared the last block of the first and second session with the first block of the second and third session, respectively, and the results are shown in Figure 4*b*. This revealed no significant difference between the last and first blocks of sessions 1 and 2, which were recorded the same day and often nearly consecutively. In contrast, there was a robust recovery in the first subblock of session 3 compared with the end of session 2 ($p < 0.05$, Wilcoxon signed rank test). Extinction of responses for this subset of cells was confirmed by significant falls in the last subblocks of session 2 and 3 ($p < 0.01$, $p < 0.05$, one-sample *t* tests).

Behavioral expression of extinction learning tracks neural extinction and requires activation of $GABA_A$ receptors in VTA

During extinction procedures in the neuronal recording experiments, animals quickly stopped licking at the spout in response to

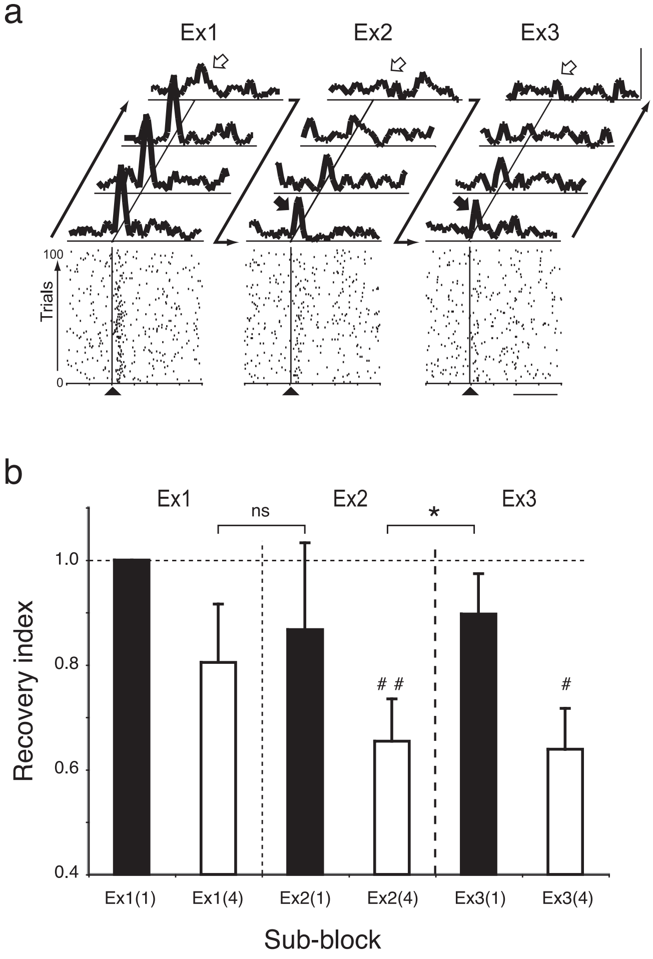

Figure 4. Spontaneous recovery of neural responses. ***a***, Single neuron. Dot rasters show action potential occurrences on all trials in three separate sessions of extinction training (Ex1, Ex2, Ex3) of a single neuron. Ex1 and Ex2 were on the same day, separated by several hours; Ex3 was performed the next day. The black triangle indicates the onset of the tone cue. Traces above the rasters show the average firing rate across subblocks of 25 trials within each session, 25 ms bins, smoothed with a three-bin moving average. Calibration: 0.5 s, 20 Hz. Reduction of the response amplitude at the end of each session (white arrowheads) indicates extinction of neural responsiveness, and spontaneous recovery is reflected in the return of the response at the beginning of the following session (black arrowheads). ***b***, Quantitative analysis of spontaneous recovery in cells tested with three sequential blocks (Ex1, Ex2, Ex3) of extinction ($n = 5$). Histogram bars show the mean + SEM recovery index (see Results) for the first and last subblocks [Ex1(1), Ex1(4), etc.] of each session. A dotted vertical line divides sessions recorded on the same day, i.e., separated by a relatively short interval, whereas a dashed vertical line indicates that Ex3 was recorded the next day. $^{*}p < 0.05$, Wilcoxon signed rank test; $^{\#}p < 0.05$, $^{\#\#}p < 0.01$, one-sample t test. $n = 6$ for all groups except Ex3(4) ($n = 5$).

cue or solenoid click. An example showing the changes in behavioral and neural response with changed reward value of the solenoid click in the solenoid-only paradigm is shown in Figure 5*a*. In this example, licking behavior ceased before the neural response to the solenoid click had completely faded. A representation of the time course of changes in licking and neural responses across the five cells tested with solenoid only is shown as an inset in Figure 5*a*. For cell activity, the extinction index was calculated as the modulation for each solenoid extinction block (calculated as the mean firing rate in the 50–175 ms postsolenoid epoch divided by the presolenoid baseline rate), normalized to the modulation for the last block of unsignaled reward. For licking behavior, the extinction index was the number of solenoid extinction trials in which licking occurred within 3 s of the solenoid, normalized to the number recorded in the last block of unsignaled reward. It is important to note that quantitative comparison of the curves shown is difficult because licking either occurs or not on a trial, whereas neural modulation is continuously graded, and there are many mechanisms by which neural activity and behavioral output could be linked by nonlinear, threshold phenomena. In the absence of knowledge about the depth of DA neural modulation required to modify behavioral output, it is therefore not possible from these data to draw conclusions about whether neural or behavioral changes "lead." All that can be said is that at the time resolution of this analysis (bins of 10 trials), neural and behavioral responding begin to drop together during extinction of responses to the well-trained solenoid click stimulus.

The behavioral response to tone cues was not quantified in the original experiments, but tests in a separate group using the same conditioning and extinction paradigms over three blocks of extinction training, shown in Figure 5*b*, confirmed the within-session reduction in licking responses during cues-only trials and spontaneous recovery at the beginning of subsequent sessions. Comparing these data with changes in neural activity across blocks of extinction sessions (Fig. 4*b*) suggests, similar to solenoid extinction, that extinction of the neural and behavioral responses track together over a broadly similar time course.

The finding that DA neurons develop inhibitory troughs in PSTHs during extinction training raises the possibility that inhibitory inputs to DA cells may become more active or be strengthened by these procedures. This raises the question whether inhibitory inputs acting on DA neurons are necessary for the expression of normal extinction behavior.

To further assess whether inhibitory inputs to the general region containing DA cells are important for the initial extinction of behavioral responding, we measured extinction rate over a single session in a group of rats with chronically implanted cannulas during reversible blockade of GABA_A receptors in the vicinity of the DA cell bodies by intracerebral infusion of bicuculline. Comparison of treated and vehicle control groups over the course of the trial subblocks within the session (Fig. 5*c*) showed that the animals treated with bicuculline extinguished significantly more slowly than the controls, so that the two groups separated steadily over trials (ANOVA with extraction of orthogonal-component trial blocks × group, linear: $F = 5.25$; $\text{df} = 1.40$; $p = 0.027$).

DA cell activity and behavioral properties of extinction can be modeled by a temporal difference algorithm that incorporates separate excitatory and inhibitory channels

To explore the implications of the experimental results for theories of extinction learning, we constructed a model using the well-characterized TD algorithm as a basis for achieving cue–reward associations. A schematic of the model, which was based on that used to explore initial learning in these DA cells, is shown in Figure 6*a*. The important additions we introduced to the basic TD algorithm were as follows: (1) parallel excitatory and inhibitory weights that are differentially affected by reward learning and extinction, and (2) different rates of spontaneous decay of excitatory and inhibitory weights. Neither the choice of a serial compound stimulus representation in the model, nor the specific use of a TD engine to achieve temporal-credit assignment were necessarily essential for our purpose; other model approaches capable of supporting such learning with prediction errors matching DA cell activity patterns might serve equally as platforms from which to study the implications of dual excitatory/inhibitory pathways.

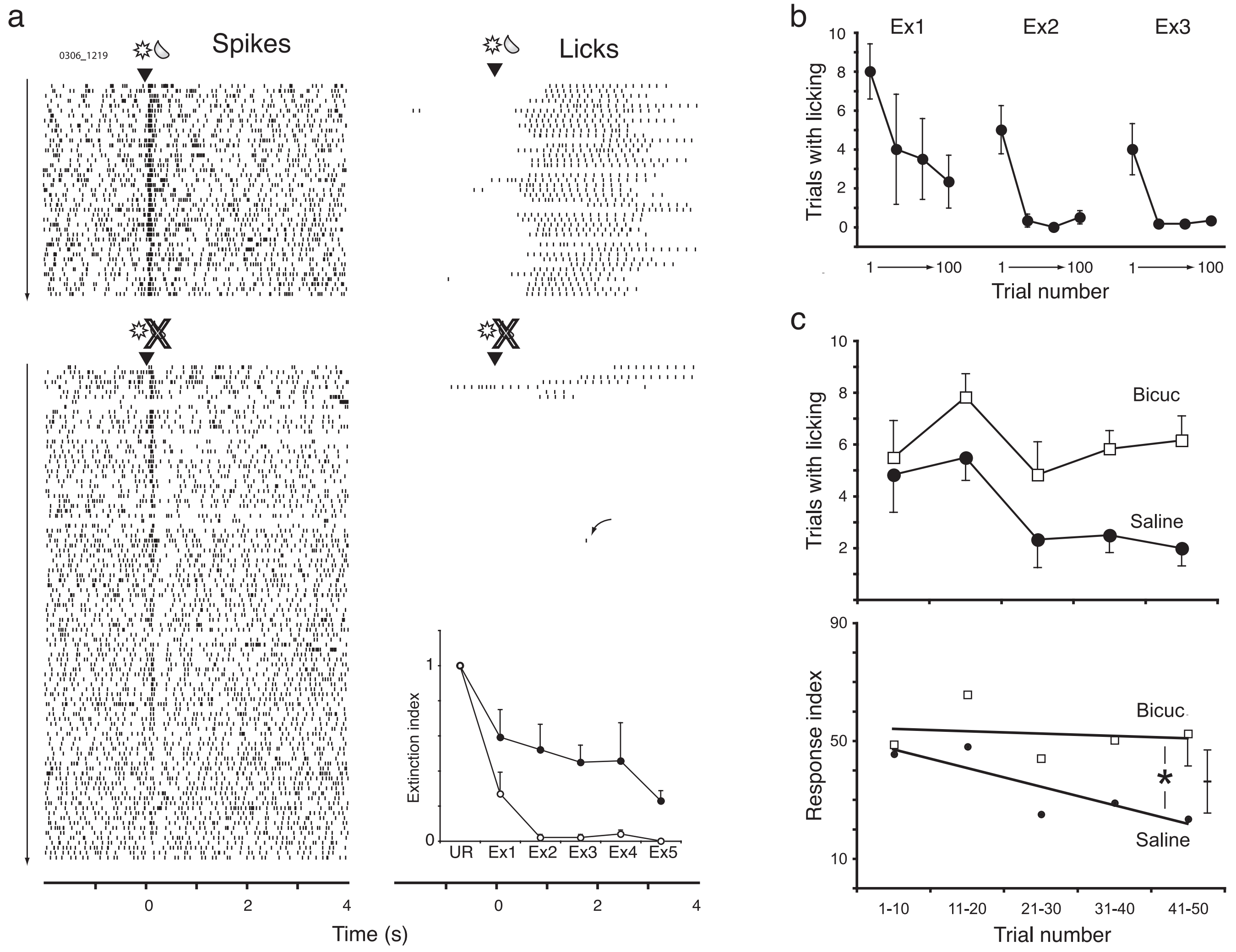

Figure 5. Behavioral extinction. ***a***, Parallel extinction of behavioral and neural responding during solenoid extinction. Rasters show example of changes in neural activity and behavior for one DA cell recording. Left rasters show spikes of the cell on each trial of unsignaled reward (top) and solenoid extinction (bottom), whereas rasters on right show all licks made at the spout over the same trials. Solenoid click time is indicated by the star symbol; a droplet symbol indicates that fluid was also delivered or, if crossed out, that fluid was not delivered. Trials are in time order with the first at the top. The curved arrow indicates the last lick made during the session. The inset graph shows mean data across all cells tested ($n = 5$). Filled and open circles show cell firing rate and licking behavior, respectively, expressed as an extinction index (see Results), in the last 10 trials of the unsignaled reward (UR) session and for five sequential blocks of 10 trials of the solenoid extinction session (Ex1–Ex5). Error bars show SEM. ***b***, Time course of extinction of behavioral responses to cue tones, recorded in a separate series of animals. Points show the mean (±SEM; $n = 6$ rats) number of trials on which any lick occurred for successive 25-trial blocks over three successive cues-only training sessions. ***c***, Effect on behavioral extinction of injection of GABA antagonist into VTA. The top graph shows mean ± SEM number of trials with licking responses across consecutive blocks of 10 trials during extinction training. Open squares, Extinction training with intracerebral bicuculline (Bicuc) infusion; filled circles, vehicle (saline). The bottom graph shows statistical analysis. The vertical axis shows the conditioned response index (arcsine transformed, in degrees, to normalize the variance). Lines show the linear regression of transformed data. The error bar on last bicuculline point shows within-subject SEM; the floating error bar indicates ±SEM across rats. *$p < 0.05$, ANOVA on the linear component comparing the two data sets overall.

The model was tested under three conditions: acquisition, extinction, and relearning. The weight changes that underlie the performance of the model were described in Materials and Methods and are schematically illustrated in exaggerated form in Figure 6*b*. In brief, as in standard TD algorithms, weights are initially set at 0. During the conditioning phase, when stimulus *S* is paired with reward *r*, positive prediction errors at the time of rewards actively strengthen excitatory input weights w^+. This increase in w^+ leads to the original learning that forms an association between the cue and the reward.

New to this model is the fact that negative prediction errors at the time of omitted rewards during extinction training actively strengthen the inhibitory input weights w^- (Fig. 6*b*). We refer to this active weight strengthening in the inhibitory input during extinction as new learning. At the same time, the negative prediction-error signal also drives excitatory weights back toward zero. We refer to this active process as unlearning. Unlearning occurs in any standard TD model but alone cannot allow for spontaneous recovery (Montague et al., 1996).

As schematically shown in Figure 6*b*, eventually the two opposing systems of weights exactly match in amplitude and cancel out, resulting in a zero prediction error at the time of the cue, as demonstrated by return to baseline. Importantly, because net weight is the result of a balance between excitatory and inhibitory weights in this model, extinction of output occurs before w^+ values return to 0, i.e., before the unlearning process is complete. Thus, some original learning is retained in the weights when extinction, as indicated by zero net output, has occurred. In other words, at this stage, there are two "memories" represented by separate weight changes in the model: the original acquired learn-

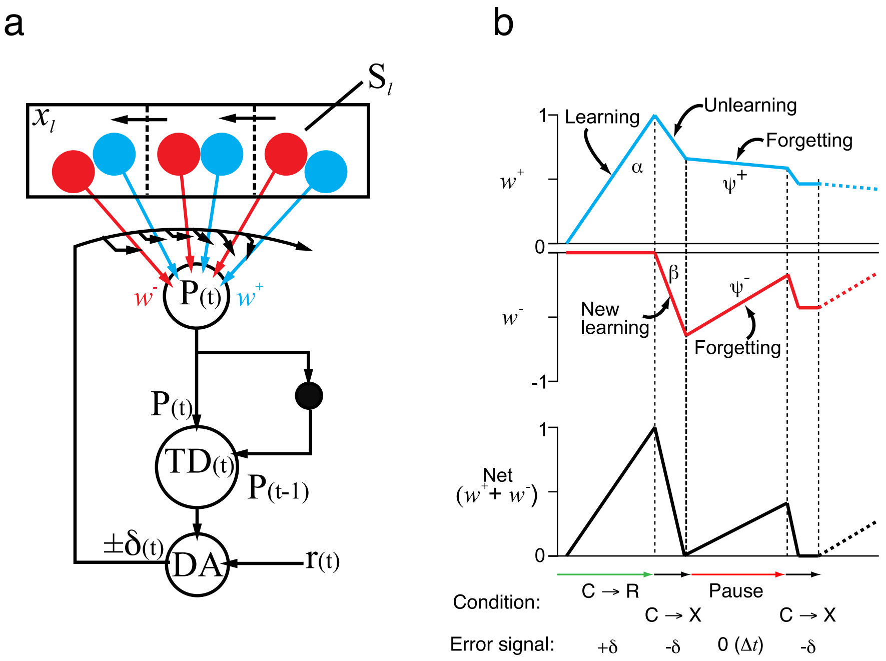

Figure 6. TD model. ***a***, Schematic diagram. Separate positive (excitatory, w^+; blue,) and negative (inhibitory, w^-; red) weights both contribute to generation of predictions (*P*) after an external sensory stimulus (S_l) represented by state vector x_l. TD, Temporal difference of predictions; r, reward; δ, prediction error, postulated to be represented in the firing of DA neurons. The prediction error feeds back to influence both positive and negative weights. ***b***, Schematic diagram demonstrating weight changes in the model during learning and extinction. Slopes have been exaggerated to emphasize important features. The lines illustrate how, during different phases, changes in positive (blue) and negative (red) weights result in net weight changes (black) that underpin conditioning, extinction, and spontaneous recovery. Positive weights are strengthened (learning) by positive prediction error signals (+δ) that occur when cues are paired with rewards (C → R), and weakened (unlearning) by negative prediction errors (−δ) generated during extinction training, when cues do not predict rewards (C → X). Rates of both changes are determined by the parameter α. Conversely, negative weights are strengthened by negative prediction error signals during extinction, at a rate set by parameter β. When there is no prediction error signal, both weights undergo decay (forgetting) as a function of time (Δ*t*), at rates determined by parameters $\psi+$ and ψ^- for positive and negative weights, respectively.

ing (positive weights) and the new extinction learning (negative weights). Finally, we incorporated a passive process we term "forgetting." In contrast to unlearning, which is driven by repeated presentation of unrewarded stimuli, forgetting in the model is the spontaneous, time-dependent decay of weights in the absence of any specific extinction stimulus.

Actual output of the model, in terms of the prediction-error signal (δ) that is equated with dopamine cell activity, is shown in Figure 7 for a range of parameter settings to demonstrate the effect of varying the ratios of learning and decay rate parameters. The three-dimensional (3D) surface plots in Figure 7*a* illustrate prediction error output at all time steps in each trial, including cue and reward responses, across all trials, whereas Figure 7*b* shows just the time step containing the cue response. Inspection of the plots reveals that varying the β/α ratio leads to relatively subtle changes in appearance, without fundamentally influencing the ability of the model to produce spontaneous recovery. This even remains true for the case where $\beta = \alpha$ (data not shown), except that in this case many more trials would be required. In contrast, the results in Figure 7 demonstrate that spontaneous recovery does not occur when $\psi^-/\psi^+ = 1$ (Fig. 7*a*,*b*, top row). A peak visible in the top left graph, where $\psi^-/\psi^+ = 1$ and $\beta/\alpha = 20$ is not spontaneous recovery. Rather, it reflects the fact that, for that combination of parameters, extinction was not yet complete at the end of the extinction session. Inequality of ψ^+ and ψ^- is thus the only key parameterization essential for spontaneous recovery.

The larger 3D plot in Figure 7*a* shows the combination of parameters that best recreated important features of learning and extinction, with a relative time course for these phases that approximated what was seen in the cell recordings. This plot used decay factors of $\psi^+ = 0.999999$ and $\psi^- = 0.9999$ for a ψ^+/ψ^- ratio of 0.999901, and learning rates of $\alpha = 0.005$ and $\beta = 0.2$ for a β/α ratio of 40. The value of α was that used in our model of acquisition by these neurons (Pan et al., 2005), whereas β was set to achieve rapid extinction as seen in the neural data. Importantly, performance during initial learning by this model is unchanged compared with simpler models designed to account only for that phase (Pan et al., 2005); responses to the cue grow with learning, whereas responses to the predicted reward decline.

After the learning phase, we exposed the model to extinction training (cue not followed by reward), which led to a decline in response to the cue. This is shown more clearly in the line plots of Figure 7*b*. Then, after a period during which no stimuli are delivered, further tests of cue delivery once again trigger a response, i.e., spontaneous recovery. This then declines once more with repeated presentations of the cue, with this sequence recurring on a second test after another pause in cue deliveries.

The time course of extinction and spontaneous recovery in the model across the three extinction test phases is presented in more detail in Figure 8*a* as the average prediction error output for subblocks of 25 trials. The relative sizes of the first and last subblocks reflect extinction, and differences between the last subblock of one session and the first of the next reflect spontaneous recovery. Comparison of the model outputs with Figures 4*b* and 5*b* confirms that the model successfully captures both the neuronal and behavioral data of the present study, as well as previously reported behavioral results (Rescorla, 2001, 2004).

Although the presence of an inhibitory channel for the model was inspired by the observation of biphasic responses in some neurons, the algorithm as presented here does not itself generate inhibitory "troughs" in the prediction-error response. This is because the action of the weights was limited to a single time step, i.e., had no latency or temporal extension, and learning stopped once excitatory and inhibitory weights were exactly balanced. Thus, the model addresses a situation in which inhibitory and excitatory weights both act over an identical time course and are exactly equivalent in effect. To produce biphasic responses in the model would simply require that weights have a latency and temporal extension, and for excitatory and inhibitory weights to have different values for those parameters.

Spontaneous recovery of behavioral responses after extinction displays several interesting and nonintuitive features that should be accounted for by any model of extinction. In particular, the level of spontaneous recovery of behavioral responses diminishes with repeated testing (Rescorla, 2004), whereas, conversely, the extent of spontaneous recovery is greater for longer than for

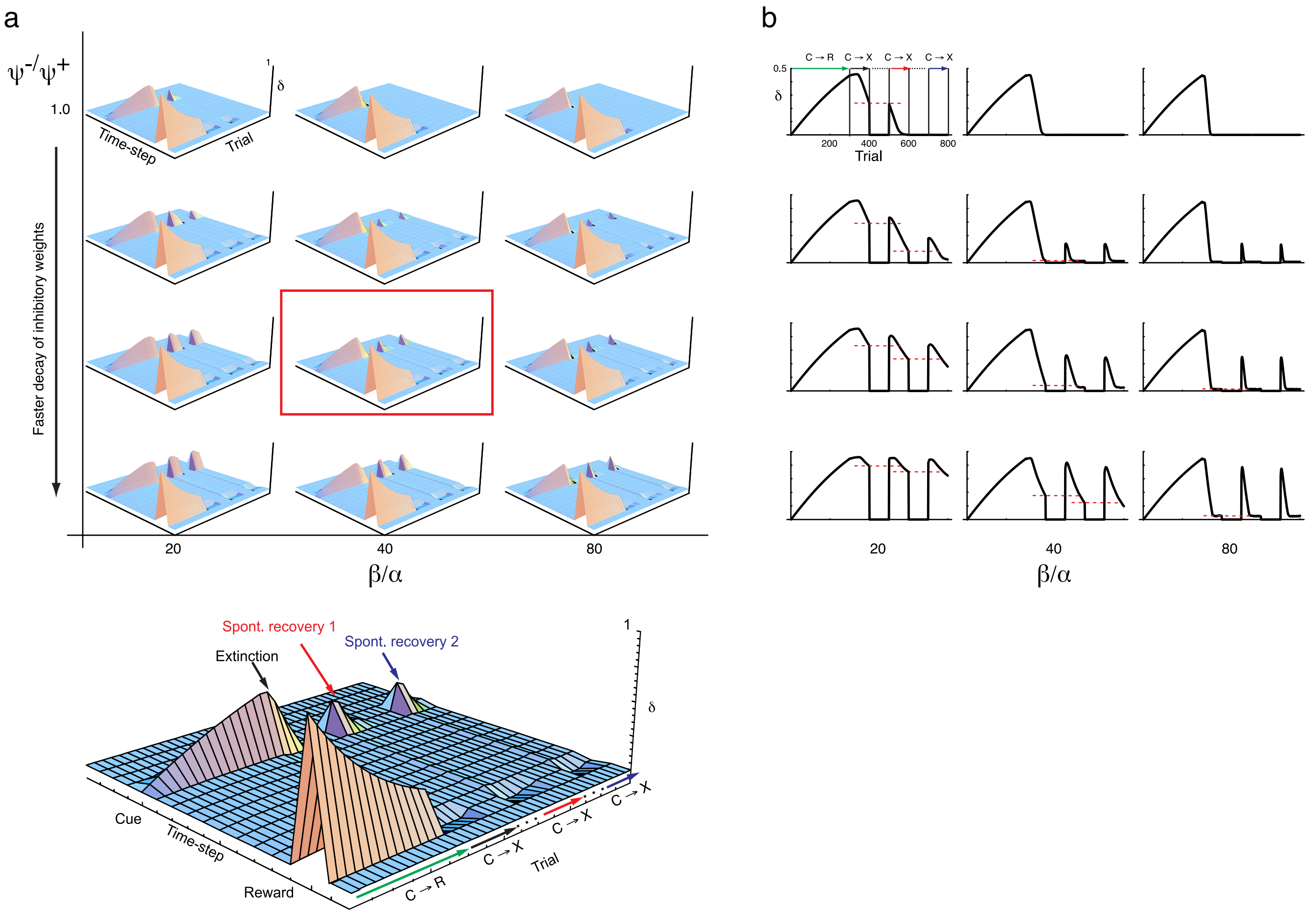

Figure 7. Prediction-error output of the TD model during learning and extinction. ***a***, Exploration of parameter space. 3D plots show prediction-error responses to cues and rewards for each time step during learning (300 trials), extinction (100 trials), and tests for spontaneous recovery (100 trials each), for different ratios of decay (ψ^-/ψ^+), and learning (β/α) parameters. For the *y*-axis, values of ψ^- range from 0.999999 to 0.9998 with $\psi^+ = 0.999999$, for ratios (top to bottom) of 1, 0.999951, 0.999901, and 0.999801. For the *x*-axis, values of β range from 0.1 to 0.4, with $\alpha = 0.005$, for ratios (left to right) of 20, 40, and 80. The outlined plot is shown enlarged below. Note that the profile of cue and reward responses during cue–reward association learning (C ⟶ R, green arrow) are the same as in the standard TD model (Pan et al., 2005). During extinction training (cues without rewards, C ⟶ X, black arrow), the cue response disappears, but after a pause (equivalent in duration to 100 trials) during which no cues are delivered, the cue response briefly reappears for initial trials when the cue is once again presented (spontaneous recovery, red arrow) but rapidly falls away. Retesting after a further pause reveals a further spontaneous recovery (blue arrow). ***b***, Cue responses across the parameter space. Plots show the prediction error across trials of learning, extinction, and tests for spontaneous recovery, at the time step of cue delivery, for each of the 3D plots in ***a***. Horizontal dashed red lines indicate the level of prediction error output obtained on the last trial of the extinction training. A prediction error signal above this level on the first trial of subsequent reexposure to cues indicates spontaneous recovery.

shorter intertest intervals (Myers and Davis, 2002). Finally, if animals are retrained after extinction, they learn more quickly than during the original training (Hull, 1943; Medina et al., 2001; Bouton and Moody, 2004). The present model generated all of these features. In Figure 7*b*, different heights of the peaks in the second and third extinction epochs demonstrate the first effect, diminishing spontaneous recovery with repeated testing. The prominence of the effect depended on an interaction of settings for the ψ and α/β ratios. Figure 8*b* shows the result of running the model with double the interval between extinction and testing for spontaneous recovery. This led to an increase in the amplitude of the spontaneously recovered prediction-error output, as predicted. Finally, the model output data in Figure 8*c* shows that, compared with the initial learning rate (first learning), the rate of relearning by the model is speeded after extinction. This is true for both relearning immediately after extinction, and relearning after a pause that allows the development of spontaneous recovery.

Discussion

Extinction is a complex phenomenon and has never been fully explained by a single mechanism. Our results indicate that dopamine neurons are engaged in specific ways during extinction of reward-mediated learning. Modeling of their activity suggests that this form of extinction can be considered as a unified mechanism involving the dynamic interaction of three components: new learning, forgetting, and unlearning.

Inhibitory responses in DA cells during extinction represent new learning

We provide the first evidence for a specific extinction-related pattern of response in DA neurons, consisting of two components. First, there was a gradual reduction in the amplitude of excitations over a similar time course to extinction of behavioral measures of cue–reward associative learning. Second, many cells developed a new inhibitory response, which either replaced the excitation or left a biphasic excitatory–inhibitory response. Several mechanisms could account for the development of these stimulus-locked inhibitions, including phasic reduction in the level of a tonic excitatory input or, more directly, enhancement of any one of the extensive inhibitory inputs to DA neurons (Tepper and Lee, 2007). Either of these mechanisms requires plasticity of inhibitory systems somewhere in the circuit that modifies DA

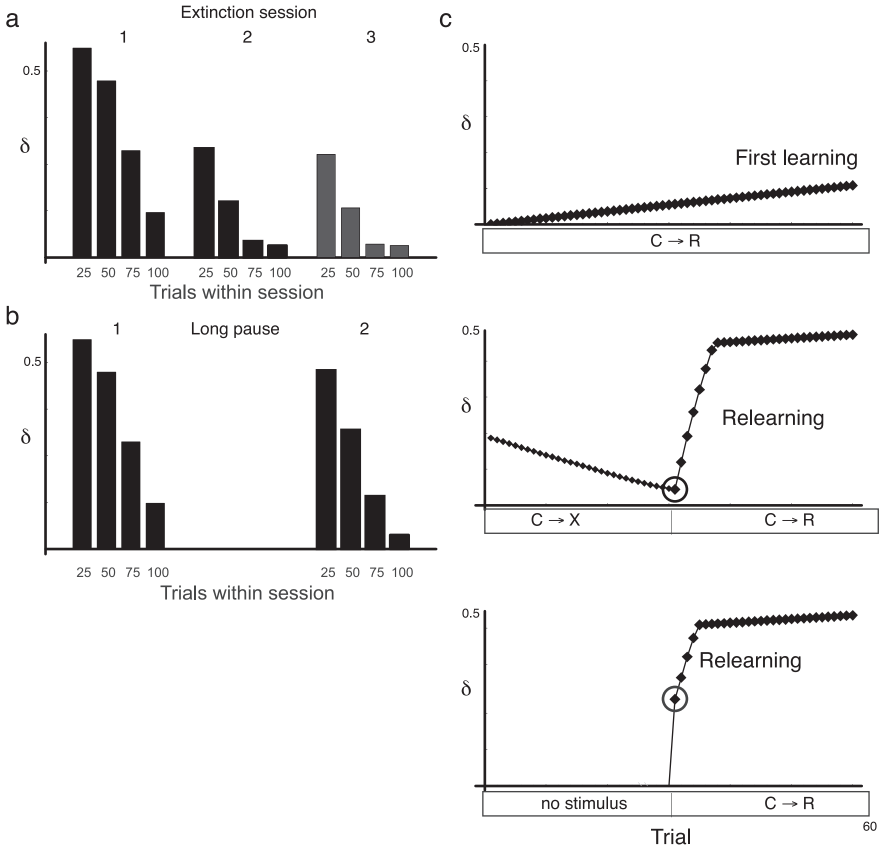

Figure 8. Patterns of model output during extinction match behavioral features of extinction learning. ***a***, Prediction error responses of the model (δ) to cue signals within and across three extinction sessions, averaged from the data shown in the enlarged 3D surface in Figure 6*a* and formatted to compare with the behavioral and neuronal data shown in Figure 2, *b* and *c*. ***b***, A longer delay between extinction sessions 1 and 2 (twice as long as in ***a***) results in enhanced spontaneous recovery at the beginning of the second session. ***c***, Speeded relearning after extinction. Points show height of the prediction error response to cue over sequential trials. Top plot shows growth in prediction-error response to cue during initial learning of cue–reward association (C ⟶ R). The middle plot shows loss of response during extinction (C ⟶ X), and faster rate of growth in prediction error response when cue–reward association is reinstigated (C ⟶ R). The circle indicates the first C ⟶ R trial, which occurs immediately after extinction, so there is no spontaneous recovery. The bottom plot shows speeded relearning after a postextinction pause, during which no stimuli are delivered. Here, relearning is superimposed on the spontaneous recovery of cue response that occurs on the first trial (circle).

neural responses in extinction, providing an example at the level of single neurons in the brain of the kind of inhibitory new learning that is often assumed to underlie behavioral extinction (Pavlov, 1927; for review, see Bouton and Moody 2004).

The concept that inhibitory processes are required for extinction is supported by the results from systemic administration of GABA antagonists (Harris and Westbrook, 1998), but the location in the brain of this effect was unclear. Our results from infusion of GABA antagonist directly into the VTA suggest that inhibitory inputs to this region are important. Although an effect on nondopaminergic outputs from the VTA cannot be ruled out, this result is consistent with other evidence pointing to an important role for inhibition at the level of DA cells in extinction processes. For instance, DA cells are the only population we are aware of that have been shown to exhibit such inhibition; studies of other brain regions involved in learning, such as cerebral cortex, thalamus, and amygdala only report reduction in excitation (Diamond and Weinberger, 1986; Komura et al., 2001; Toyomitsu et al., 2002). More directly, DA cells in monkey have also been found to display inhibitory cue responses in other situations, not explicitly related to extinction, where the cues are nevertheless associated with failure of subsequent reward delivery. Thus, DA cells are inhibited by cues that always predict the absence of reward (Tobler et al., 2003). Similarly, inhibitions are seen in cases where the cells develop unexpected "erroneous" responses in classical learning paradigms (Waelti et al., 2001). Finally, inhibitory components are prominent when DA cell responses generalize to incorrect stimuli during learning (Mirenowicz and Schultz, 1996). Furthermore, similar to the present findings, these inhibitory responses are often part of excitation–inhibition sequences.

Parallel excitatory and inhibitory inputs to DA cells may underpin learning and extinction

Together, these data are consistent with the idea that DA neurons are a site of convergence of excitatory inputs carrying information about sensory cues [e.g., from the pedunculopontine tegmental nucleus (Pan and Hyland, 2005; Kobayashi and Okada, 2007) and superior colliculus (Comoli et al., 2003)], and inhibitory inputs that superimpose a suppressive control of responses to cues depending on their salience (Schultz, 1998). Potential sources of inhibition include both striatal projection neurons and local inhibitory interneurons (Tepper and Lee, 2007). These inhibitory structures may mediate inhibition regulated by the lateral habenula (Ji and Shepard, 2007), which has been shown recently to have responses consistent with negative reward signaling (Matsumoto and Hikosaka, 2007). If local inhibitory neurons are involved, then reciprocal patterns of activity in non-DA cells would be expected. Few cells with high firing rates and narrow waveforms typical of inhibitory interneurons were fully tested in the present study. However, the potential for such responding is suggested by one cell, presented in supplemental material (available at www.jneurosci.org), that showed an inhibitory response to conditioned cues, but developed an excitatory response to the same cues during extinction.

In addition to inhibitory responses to conditioned cues, DA cells in monkey (Hollerman and Schultz, 1998) and rat (Pan et al., 2005) (supplemental material, available at www.jneurosci.org) also show suppression of activity around the time a reward is expected, but omitted. These responses provide further confirmation of the importance of inhibitory inputs in regulating DA error signaling, and might be involved in generating the earlier cue-related inhibitions we describe here. The inhibition of DA cells at this time has been proposed to represent a negative prediction error for learning (Hollerman and Schultz, 1998; Schultz, 1998; Satoh et al., 2003), which could drive plastic changes in brain circuits that respond to earlier sensory cues. Such new learning would account for the changes seen in DA cell responses to cues during extinction seen in the present study. The strengthened inhibitory input to cues during extinction would then serve to suppress responses to signals, which have been learned to not be associated with reward.

To examine the functional implications of the proposal that separate, modifiable excitatory and inhibitory pathways converging on DA cells are important for extinction, we examined the effect of incorporating inhibitory processes into a TD model. A key feature of the model is that excitatory and inhibitory weights can be modulated in opposite directions during learning and extinction, reflecting the observation that in DA cells, inhibitory responses arose during extinction, whereas excitatory responses declined. There is some evidence that such differential and opposite effects can be evoked at the synaptic level. Substance P can cause excitatory and inhibitory synapses to be modulated in opposite directions by a common stimulus (Parker and Grillner, 1999), and cholinergic forebrain systems can simultaneously potentiate excitatory and depotentiate inhibitory inputs to cortical neurons (Froemke et al., 2007).

We found that the dual excitatory–inhibitory model was able to recreate extinction of the learned prediction-error outputs, as well as modeling the patterns of prediction-error output during learning in the same way as achieved previously by standard TD (Montague et al., 1996; Schultz et al., 1997; Pan et al., 2005). Standard TD models can produce simple elimination of learned responses through unlearning (Montague et al., 1996). What is new for this model is that it also exhibits spontaneous recovery, a classical observation of behavioral extinction that we saw in the DA neuronal activity. Furthermore, the model also exhibits other characteristics behavioral extinction, such as speeded relearning (Rescorla, 2004) and improved recovery with longer intertest intervals (Myers and Davis, 2002). It will be of interest to determine whether DA cells responses follow these predicted patterns.

To achieve spontaneous recovery, a key requirement was differential spontaneous decay rates for excitatory and inhibitory weights. An interesting proposition arising from the model is therefore that such differences occur in the brain. In behavioral studies, learning related to inhibiting a response has been found to have a faster decay rate than learning to emit one (Hendersen, 1978), but this does not necessarily relate to inhibition/excitation at the neural level. Furthermore, it is not necessarily the case that differences in decay rate link to sign of synaptic action or the transmitter used. For instance, recency of plastic change, rather than properties of the synapse per se, may be a critical factor. Learning involves several phases including protein-expression dependent mechanisms for consolidation (Miyashita et al., 2007). If newer learning is less consolidated, it may be more liable to decay. Indeed, it is well established in the behavioral literature that for two learned associations of unequal age, the older association will decay less rapidly than the newer (Wheeler et al., 2004; Wixted, 2004). In extinction learning, extinction procedures follow the initial training, and thus represent newer, potentially more labile learning. Whichever is the critical factor, differential decay rates for plasticity in different synapses or circuits at the physiological level are a testable prediction of the model.

Extinction as a tripartite process involving learning, forgetting, and unlearning

The neural and modeling results described above point to a key role for plasticity in inhibitory circuits during extinction. Furthermore, the model predicts differential spontaneous decay (or forgetting) rates for inhibitory and excitatory weights. The model thus formalizes the hypothesis, widely held since the study by Pavlov (1927), that spontaneous recovery can be accounted for by new, extinction-induced inhibitory learning that suppresses the old, combined with faster decay of the extinction learning (Bouton and Moody, 2004; Delamater, 2004; Rescorla, 2004). However, although new learning and forgetting processes may be essential for the property of spontaneous recovery, there could still be a role for unlearning (the active erasure of learning) (Rescorla and Wagner, 1972; McClelland and Rumelhart, 1985). Recent behavioral studies point to specific situations in which unlearning may be significant (Medina et al., 2002; Rescorla, 2004; Myers et al., 2006). Active unlearning is intrinsic to TD models, with the relative contribution it makes dependent on choice of parameters; in the present model, it played a very minor role. Nevertheless, appropriate settings may model these specific situations. Extinction may involve contributions from all three mechanisms, with the relative contribution of new learning, forgetting, and unlearning processes varying depending on the situation (Mauk and Ohyama, 2004).

References

Aebischer P, Schultz W (1984) The activity of pars compacta neurons of the monkey substantia nigra is depressed by apomorphine. Neurosci Lett 50:25–29.

Bouton ME, Moody EW (2004) Memory processes in classical conditioning. Neurosci Biobehav Rev 28:663–674.

Chiodo LA, Bannon MJ, Grace AA, Roth RH, Bunney BS (1984) Evidence for the absence of impulse-regulating somatodendritic and synthesis-modulating nerve terminal autoreceptors on subpopulations of mesocortical dopamine neurons. Neuroscience 12:1–16.

Comoli E, Coizet V, Boyes J, Bolam JP, Canteras NS, Quirk RH, Overton PG, Redgrave P (2003) A direct projection from superior colliculus to substantia nigra for detecting salient visual events. Nat Neurosci 6:974–980.

Delamater AR (2004) Experimental extinction in pavlovian conditioning: behavioural and neuroscience perspectives. Q J Exp Psychol B 57:97–132.

Diamond DM, Weinberger NM (1986) Classical conditioning rapidly induces specific changes in frequency receptive fields of single neurons in secondary and ventral ectosylvian auditory cortical fields. Brain Res 372:357–360.

Froemke RC, Merzenich MM, Schreiner CE (2007) A synaptic memory trace for cortical receptive field plasticity. Nature 450:425–429.

Harris JA, Westbrook RF (1998) Evidence that GABA transmission mediates context-specific extinction of learned fear. Psychopharmacology (Berl) 140:105–115.

Heil P (1997) Auditory cortical onset responses revisited. I. First-spike timing. J Neurophysiol 77:2616–2641.

Hendersen RW (1978) Forgetting of conditioned fear inhibition. Learn Motiv 9:16–30.

Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. Nat Neurosci 1:304–309.

Hull CL (1943) Principles of behavior. New York: Appleton-Century-Crofts.

Hyland BI, Reynolds JN, Hay J, Perk CG, Miller R (2002) Firing modes of midbrain dopamine cells in the freely moving rat. Neuroscience 114:475–492.

Ji H, Shepard PD (2007) Lateral habenula stimulation inhibits rat midbrain dopamine neurons through a GABA(A) receptor-mediated mechanism. J Neurosci 27:6923–6930.

Klopf AH (1988) A neuronal model of classical conditioning. Psychobiology 16:85–125.

Kobayashi Y, Okada K (2007) Reward prediction error computation in the pedunculopontine tegmental nucleus neurons. Ann N Y Acad Sci 1104:310–323.

Komura Y, Tamura R, Uwano T, Nishijo H, Kaga K, Ono T (2001) Retrospective and prospective coding for predicted reward in the sensory thalamus. Nature 412:546–549.

Konorski J (1948) Conditioned reflexes and neuron organisation. New York: Cambridge UP.

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

Matsumoto M, Hikosaka O (2007) Lateral habenula as a source of negative reward signals in dopamine neurons. Nature 447:1111–1115.

Mauk MD, Ohyama T (2004) Extinction as new learning versus unlearning: considerations from a computer simulation of the cerebellum. Learn Mem 11:566–571.

McClelland JL, Rumelhart DE (1985) Distributed memory and the representation of general and specific information. J Exp Psychol Gen 114:159–197.

Medina JF, Garcia KS, Mauk MD (2001) A mechanism for savings in the cerebellum. J Neurosci 21:4081–4089.

Medina JF, Nores WL, Mauk MD (2002) Inhibition of climbing fibres is a signal for the extinction of conditioned eyelid responses. Nature 416:330–333.

Mirenowicz J, Schultz W (1996) Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. Nature 379:449–451.

Miyashita T, Kubik S, Lewandowski G, Guzowski JF (2007) Networks of neurons, networks of genes: an integrated view of memory consolidation. Neurobiol Learn Mem 89:269–284.

Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 16:1936–1947.

Myers KM, Davis M (2002) Behavioral and neural analysis of extinction. Neuron 36:567–584.

Myers KM, Ressler KJ, Davis M (2006) Different mechanisms of fear extinction dependent on length of time since fear acquisition. Learn Mem 13:216–223.

Pan WX, Hyland BI (2005) Pedunculopontine tegmental nucleus controls conditioned responses of midbrain dopamine neurons in behaving rats. J Neurosci 25:4725–4732.

Pan WX, Schmidt R, Wickens JR, Hyland BI (2005) Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. J Neurosci 25:6235–6242.

Parker D, Grillner S (1999) Activity-dependent metaplasticity of inhibitory and excitatory synaptic transmission in the lamprey spinal cord locomotor network. J Neurosci 19:1647–1656.

Pavlov IP (1927) Conditioned reflexes. Oxford: Oxford UP.

Paxinos G, Watson C (1997) The rat brain in stereotaxic coordinates, Ed 3. London: Academic.

Rescorla RA (2001) Retraining of extinguished pavlovian stimuli. J Exp Psychol Anim Behav Process 27:115–124.

Rescorla RA (2004) Spontaneous recovery. Learn Mem 11:501–509.

Rescorla RA, Wagner AR (1972) A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In: Classical conditioning II: current research and theory (Black AH, Prokasy WF, eds), pp 64–99. New York: Appleton-Century-Crofts.

Roesch MR, Calu DJ, Schoenbaum G (2007) Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat Neurosci 10:1615–1624.

Satoh T, Nakai S, Sato T, Kimura M (2003) Correlated coding of motivation and outcome of decision by dopamine neurons. J Neurosci 23:9913–9923.

Schultz W (1998) Predictive reward signal of dopamine neurons. J Neurophysiol 80:1–27.

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

Sutton RS (1988) Learning to predict by the methods of temporal differences. Mach Learn 3:9–44.

Tepper JM, Lee CR (2007) GABAergic control of substantia nigra dopaminergic neurons. In: GABA and the basal ganglia: from molecules to systems (Tepper JM, Abercrombie ED, Bolam JP, eds). Amsterdam: Elsevier.

Tobler PN, Dickinson A, Schultz W (2003) Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. J Neurosci 23:10402–10410.

Toyomitsu Y, Nishijo H, Uwano T, Kuratsu J, Ono T (2002) Neuronal responses of the rat amygdala during extinction and reassociation learning in elementary and configural associative tasks. Eur J Neurosci 15:753–768.

Waelti P, Dickinson A, Schultz W (2001) Dopamine responses comply with basic assumptions of formal learning theory. Nature 412:43–48.

Wheeler DS, Stout SC, Miller RR (2004) Interaction of retention interval with CS-preexposure and extinction treatments: symmetry with respect to primacy. Learn Behav 32:335–347.

Wixted JT (2004) The psychology and neuroscience of forgetting. Annu Rev Psychol 55:235–269.