nature neuroscience

A causal link between prediction errors, dopamine neurons and learning

Elizabeth E Steinberg[1,2,11], Ronald Keiflin[1,11], Josiah R Boivin[1,2], Ilana B Witten[3,4], Karl Deisseroth[5–8] & Patricia H Janak[1,2,9,10]

Situations in which rewards are unexpectedly obtained or withheld represent opportunities for new learning. Often, this learning includes identifying cues that predict reward availability. Unexpected rewards strongly activate midbrain dopamine neurons. This phasic signal is proposed to support learning about antecedent cues by signaling discrepancies between actual and expected outcomes, termed a reward prediction error. However, it is unknown whether dopamine neuron prediction error signaling and cue-reward learning are causally linked. To test this hypothesis, we manipulated dopamine neuron activity in rats in two behavioral procedures, associative blocking and extinction, that illustrate the essential function of prediction errors in learning. We observed that optogenetic activation of dopamine neurons concurrent with reward delivery, mimicking a prediction error, was sufficient to cause long-lasting increases in cue-elicited reward-seeking behavior. Our findings establish a causal role for temporally precise dopamine neuron signaling in cue-reward learning, bridging a critical gap between experimental evidence and influential theoretical frameworks.

Much of the behavior of humans and other animals is directed toward seeking out rewards. Learning to identify environmental cues that provide information about where and when natural rewards can be obtained is an adaptive process that allows this behavior to be distributed efficiently. Theories of associative learning have long recognized that simply pairing a cue with reward is not sufficient for learning to occur. In addition to contiguity between two events, learning also requires the subject to detect a discrepancy between an expected reward and the reward that is actually obtained[1].

This discrepancy, or reward prediction error (RPE), acts as a teaching signal that is used to correct inaccurate predictions. Presentation of unpredicted reward or reward that is better than expected generates a positive prediction error and strengthens cue-reward associations. Presentation of a perfectly predicted reward does not generate a prediction error and fails to support new learning. Conversely, omission of a predicted outcome generates a negative prediction error and leads to extinction of conditioned behavior. The error correction principle figures prominently in psychological and computational models of associative learning[1–6], but the neural bases of this influential concept have not yet been definitively demonstrated.

In vivo electrophysiological recordings in non-human primates and rodents have shown that putative dopamine neurons in the ventral tegmental area (VTA) and the substantia nigra pars compacta respond to natural rewards such as palatable food[7–9]. Notably, the sign and magnitude of the dopamine neuron response is modulated by the degree to which the reward is expected. Surprising or unexpected rewards elicit strong increases in firing rate, whereas anticipated rewards produce little or no change[8,10,11]. Conversely, when an expected reward fails to materialize, neural activity is depressed below baseline[8–10]. Reward-evoked dopamine release at terminal regions *in vivo* is also more pronounced when rewards are unexpected[12]. On the basis of this parallel between RPE and dopamine responses, a current hypothesis suggests that dopamine neuron activity at the time of reward delivery acts as a teaching signal and causes learning about antecedent cues[2–4]. This conception is further supported by the observation that dopamine neurons are strongly activated by primary rewards before cue-reward associations are well learned. As learning progresses and behavioral performance nears asymptote, the magnitude of dopamine neuron activation elicited by reward delivery progressively wanes[7,10].

Although the correlative evidence linking reward-evoked dopamine neuron activity with learning is compelling, little causal evidence exists to support this hypothesis. Previous studies that attempted to address the role of prediction errors and phasic dopamine neuron activity in learning employed pharmacological tools, such as targeted inactivation of the VTA[13] or administration of dopamine receptor antagonists[14] or indirect agonists[15]. Such studies suffer from the limitation that pharmacological agents alter the activity of neurons over long timescales and therefore cannot determine the contribution of specific patterns of dopamine neuron activity to behavior. Genetic manipulations that chronically alter the actions of dopamine neurons by reducing or eliminating the ability of dopamine neurons to

[1]Ernest Gallo Clinic and Research Center, University of California, San Francisco, California, USA. [2]Graduate program in Neuroscience, University of California, San Francisco, California, USA. [3]Princeton Neuroscience Institute, Princeton University, Princeton, New Jersey, USA. [4]Department of Psychology, Princeton University, Princeton, New Jersey, USA. [5]Department of Bioengineering, Stanford University, Stanford, California, USA. [6]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, California, USA. [7]Howard Hughes Medical Institute, Stanford University, Stanford, California, USA. [8]Crack the Neural Code Program, Stanford University, Stanford, California, USA. [9]Department of Neurology, University of California, San Francisco, California, USA. [10]Wheeler Center for the Neurobiology of Addiction, University of California, San Francisco, California, USA. [11]These authors contributed equally to this work. Correspondence should be addressed to P.H.J. (pjanak@gallo.ucsf.edu).

Received 15 January; accepted 2 May; published online 26 May 2013; doi:10.1038/nn.3413

fire in bursts[16,17] do alter learning, but suffer from similar problems, as the effect of dopamine neuron activity during specific behavioral events (such as reward delivery) cannot be evaluated. Other studies circumvented these issues by using optogenetic tools that permit temporally precise control of dopamine neuron activity; however, these studies failed to utilize behavioral tasks that explicitly manipulate reward expectation[18–21], involve natural rewards[20,21] or are suitable for assessing cue-reward learning[19]. Thus, despite the prevalence and influence of the hypothesis that RPE signaling by dopamine neurons drives associative cue-reward learning, a direct link between the two has yet to be established.

To address this unresolved issue, we capitalized on the ability to selectively control the activity of dopamine neurons in the awake, behaving rat with temporally precise and neuron-specific optogenetic tools[21–23] to simulate naturally occurring dopamine signals. We sought to determine whether activation of dopamine neurons in the VTA timed with the delivery of an expected reward would mimic a RPE and drive cue-reward learning using two distinct behavioral procedures.

First, we employed blocking, the associative phenomenon that best demonstrates the role of prediction errors in learning[24–26]. In a blocking procedure, the association between a cue and a reward is prevented (or blocked) if another cue present in the environment at the same time already reliably signals reward delivery[27]. It is generally argued that the absence of an RPE, supposedly encoded by the reduced or absent phasic dopamine response to the reward, prevents further learning about the redundant cue[4,28]. We reasoned that artificial VTA dopamine neuron activation paired with reward delivery would mimic a positive prediction error and facilitate learning about the redundant cue. Next, we tested the role of dopamine neuron activation during extinction learning. Extinction refers to the observed decrease in conditioned responding that results from the reduction or omission of an expected reward. The negative prediction error, which is supposedly encoded by a pause in dopamine neuron firing, is proposed to induce extinction of behavioral responding[4,29]. We reasoned that artificial VTA dopamine neuron activation timed to coincide with the reduced or omitted reward would interfere with extinction learning. In both procedures, optogenetic activation of dopamine neurons at the time of expected reward delivery affected learning in a manner that was consistent with the hypothesis that dopamine neuron prediction error signaling drives associative learning.

RESULTS

Demonstration of associative blocking

The blocking procedure provides an illustration of the essential role of RPEs in associative learning. Consider two cues (for example, a tone and a light) presented simultaneously (in compound) and followed by reward delivery. It has been shown that conditioning to one element of the compound is reduced (or blocked) if the other element has already been established as a reliable predictor of the reward[24–27]. In other words, despite consistent pairing between a cue and reward, the absence of a prediction error prevents learning about the redundant cue. Consistent with the idea that dopamine neurons encode prediction errors, putative dopamine neurons recorded *in vivo* exhibit little to no reward-evoked responses in a blocking procedure[28]. The lack of dopamine neuron activity, combined with a failure to learn in the blocking procedure, is considered to be a key piece of evidence (albeit correlative) linking dopamine RPE signals to learning. On the basis of this evidence, we determined that the blocking procedure would provide an ideal environment in which to test the hypothesis that RPE signaling by dopamine neurons can drive learning. According to this hypothesis, artificially activating dopamine neurons during reward delivery in the blocking condition, when dopamine neurons normally do not fire, should mimic a naturally occurring prediction error signal and allow subjects to learn about the otherwise blocked cue.

We first examined associative blocking of reward-seeking (**Fig. 1**) using parameters suitable for subsequent optogenetic neural manipulation. Two groups of rats were initially trained to respond for a liquid sucrose reward (unconditioned stimulus) during an auditory cue in a single cue training phase. Subsequently, a combined auditory and visual cue was presented in a compound training phase and the identical sucrose unconditioned stimulus was delivered. For subjects assigned to the blocking group, the same auditory cue was presented during single and compound phases, whereas distinct auditory cues were used for control group subjects (**Fig. 1a**); in both phases,

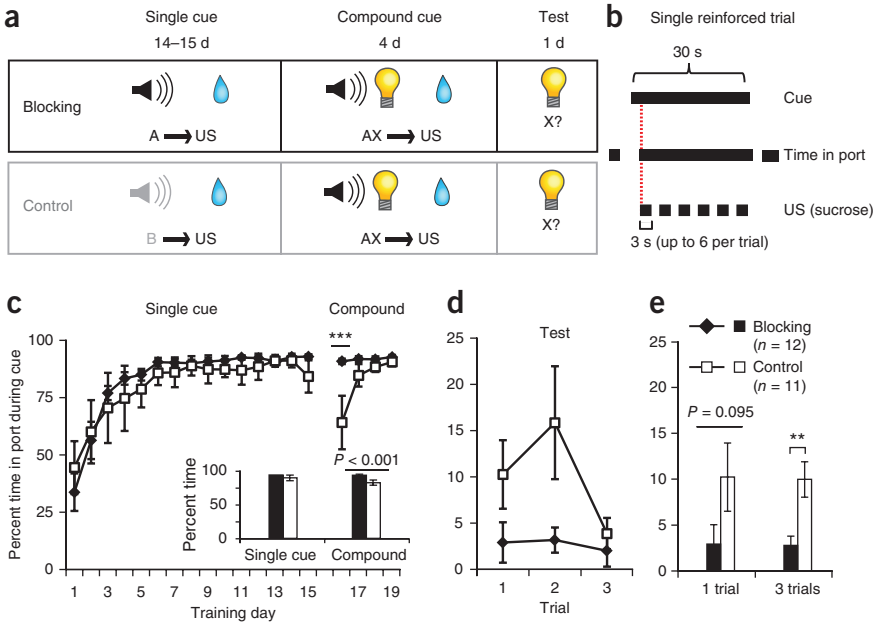

Figure 1 Behavioral demonstration of the blocking effect. (**a**) Experimental design of the blocking task. A, cue A; X, cue X; AX, compound presentation of cues A and X; US, unconditioned stimulus. (**b**) During reinforced trials, sucrose delivery was contingent on reward port entry during the 30-s cue. After entry, sucrose was delivered for 3 s, followed by a 2-s timeout. Up to six sucrose rewards could be earned per trial, depending on the rats' behavior. (**c**) Performance across all single cue and compound training sessions. Inset, mean performance among groups over the last 4 d of single-cue training did not differ; controls showed reduced behavior during compound training (***$P < 0.001$). (**d**) Performance during visual cue test. The blocking group exhibited reduced responding to the cue at test, relative to controls (main effect of group, $P = 0.003$; group × trial interaction, $P = 0.286$). (**e**) Visual cue test performance for the first trial and the average of all three trials. The blocking group showed reduced cue responding for the three-trial measure (**$P = 0.003$), but were not different on the first trial ($P = 0.095$). Data are presented as means and error bars represent s.e.m.

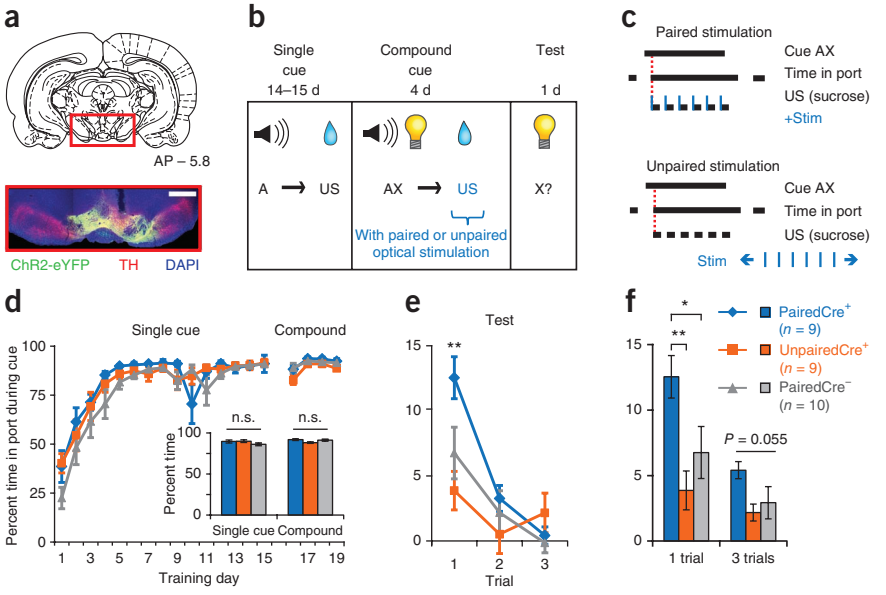

Figure 2 Dopamine neuron stimulation drives new learning. (**a**) Example histology from a *Th-cre*$^+$ rat injected with a Cre-dependent ChR2-containing virus. Vertical track indicates optical fiber placement above VTA. Scale bar represents 1 mm. (**b**) Experimental design for blocking task with optogenetics. All groups received identical behavioral training according to the blocking group design shown in **Figure 1a**. (**c**) Optical stimulation (1-s train, 5-ms pulse, 20 Hz, 473 nm) was synchronized with sucrose delivery in Paired (Cre$^+$ and Cre$^-$), but not Unpaired (Cre$^+$), groups during compound training. (**d**) Performance across all single cue and compound training sessions. Inset, no group differences were observed over the last 4 d of single cue training or during compound training. (**e**) Performance during visual cue test. The PairedCre$^+$ group exhibited increased responding to the cue relative to both control groups at test on the first trial (**$P < 0.005$). (**f**) Visual cue test performance for the first trial and all three trials averaged. The PairedCre$^+$ group exhibited increased cue responding relative to controls for the one-trial measure (PairedCre$^+$ versus UnpairedCre$^+$, **$P = 0.005$; PairedCre$^+$ versus PairedCre$^-$, *$P = 0.025$; PairedCre$^-$ versus UnpairedCre$^+$, $P = 0.26$); there was a trend for a group effect for the three-trial average (main effect of group, $P = 0.055$). Data are presented as means and error bars represent s.e.m.

unconditioned stimulus delivery was contingent on the rat's presence in the reward port during the cue (**Fig. 1b**). Thus, the critical difference between experimental groups is the predictability of the unconditioned stimulus during the compound phase; because of its prior association with the previously trained auditory cue, the unconditioned stimulus is expected for the blocking group, whereas, for the control group, its occurrence is unexpected. We measured conditioned responding as the amount of time spent in the reward port during the cue, normalized to an immediately preceding pre-cue period of equal length. Both groups showed equivalently high levels of conditioned behavior at the end of the single cue phase (two-way repeated-measures ANOVA, no effect of group or group × day interaction, all P values > 0.05), but differed in their performance when the compound cue was introduced (two-way repeated-measures ANOVA, main effect of group, $F_{1,21} = 21.15$, $P < 0.001$; group × day interaction, $F_{3,63} = 11.63$, $P < 0.001$), consistent with the fact that the association between the compound cue and unconditioned stimulus had to be learned by the control group (**Fig. 1c**).

To determine whether learning about the visual cue introduced during compound training was affected by the predictability of reward, we assessed conditioned responding to unreinforced presentations of the visual cue alone 1 d later. Conditioned responding was reduced in the blocking group as compared with controls (two-way repeated-measures ANOVA, main effect of group, $F_{1,21} = 11.27$, $P = 0.003$, no group × trial interaction, $F_{2,42} = 1.29$, $P = 0.286$; **Fig. 1d,e**), indicating that new learning about preceding environmental cues occurs after unpredicted, but not predicted, reward in this procedure, consistent with previous findings[28,30].

Reward-paired dopamine neuron activation drives learning

Putative dopamine neurons recorded in monkeys are strongly activated by unexpected reward, but fail to respond to the same reward if it is fully predicted[10,11], including when delivered in a blocking condition[28]. The close correspondence between dopamine neural activity and behavioral evidence of learning in this task suggests that positive RPEs caused by unexpected reward delivery activate dopamine neurons and lead to learning observed under control conditions. To test this hypothesis, we optogenetically activated VTA dopamine neurons at the time of unconditioned stimulus delivery on compound trials in our blocking task to drive learning under conditions in which learning normally does not occur. We used parameters that we have previously established elicit robust, time-locked activation of dopamine neurons and neurotransmitter release in anesthetized animals or *in vitro* preparations[21]. We predicted that phasic dopamine neuron activation delivered coincidently with fully-predicted reward would be sufficient to cause new learning about preceding cues.

Female transgenic rats expressing Cre recombinase under the control of the *tyrosine hydroxylase* (*Th*) promoter (*Th-cre*$^+$ rats) and their wild-type littermates (*Th-cre*$^-$ rats) were used to gain selective control of dopamine neuron activity as described previously[21]. *Th-cre*$^+$ and *Th-cre*$^-$ littermates received identical injections of a Cre-dependent virus expressing channelrhodopsin-2 (ChR2) in the VTA; chronic optical fiber implants were targeted dorsal to this region to allow for selective unilateral optogenetic dopamine neuron activation (**Fig. 2a** and **Supplementary Fig. 1**). Three groups of rats were trained under conditions that normally result in blocked learning to the light cue (cue X; **Fig. 2b**). The behavioral performance of an experimental group (PairedCre$^+$) consisting of *Th-cre*$^+$ rats that received optical stimulation (1-s train, 5-ms pulse, 20 Hz) paired with the unconditioned stimulus during compound training (see Online Methods) was compared to the performance of two control groups that received identical training, but differed either in genotype (PairedCre$^-$) or the time at which optical stimulation was delivered (UnpairedCre$^+$, optical stimulation during the intertrial interval, ITI; **Fig. 2c**). Groups performed equivalently during single cue and compound training (**Fig. 2d**), suggesting that all rats learned the task and that the optical stimulation delivered during compound training did not disrupt ongoing behavior (two-way repeated-measures ANOVA revealed no significant effect of group or group × day interaction, all P values > 0.111).

The critical comparison among groups occurred when the visual cue introduced during compound training was tested alone in an

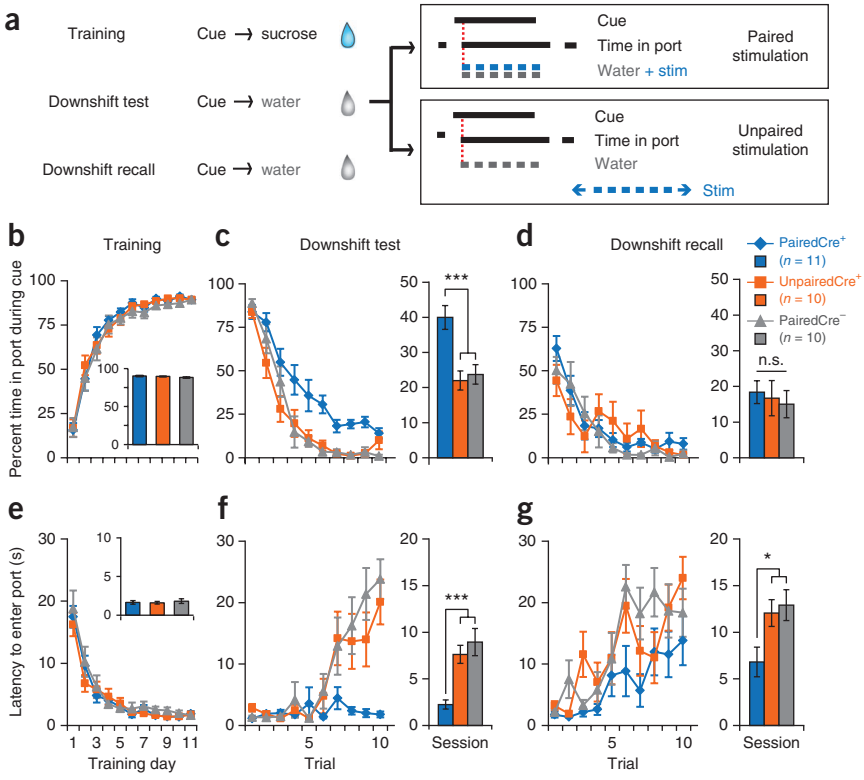

Figure 3 Dopamine neuron stimulation attenuates behavioral decrements associated with a downshift in reward value. (**a**) Experimental design for reward downshift experiment. Optical stimulation (3-s train, 5-ms pulse, 20 Hz, 473 nm) was either paired with the water reward (PairedCre$^+$ and PairedCre$^-$ groups) or explicitly unpaired (UnpairedCre$^+$) during the downshift test. (**b**) Percent time in port during the cue across training sessions. Inset, no difference in average performance during the last two training sessions. (**c**) Percent time in port during the cue for the downshift test. Data are displayed for single trials (left) and as a session average (right). PairedCre$^+$ rats exhibited increased time in port compared with controls (PairedCre$^+$ versus UnpairedCre$^+$, $^{***}P < 0.001$; PairedCre$^+$ versus PairedCre$^-$, $^{***}P < 0.001$; PairedCre$^-$ versus UnpairedCre$^+$, $P = 0.691$). (**d**) Percent time in port during the cue for downshift recall. Data are displayed for single trials (left) and as a session average (right). There were no group differences during this phase (two-way repeated-measures ANOVA, main effect of group, $P = 0.835$). (**e**) Latency to enter the reward port after cue onset. Inset, no group differences during last two training sessions. (**f**) Data are presented as in **c**, but for latency. PairedCre$^+$ rats responded faster to the cue compared with controls during the downshift test (PairedCre$^+$ versus UnpairedCre$^+$, $^{***}P < 0.001$; PairedCre$^+$ versus PairedCre$^-$, $^{***}P < 0.001$; PairedCre$^-$ versus UnpairedCre$^+$, $P = 0.375$). (**g**) Data are presented as in **d**, but for latency. PairedCre$^+$ rats responded faster to the cue than controls during downshift recall (PairedCre$^+$ versus UnpairedCre$^+$, $P = 0.024$; PairedCre$^+$ versus PairedCre$^-$, $P = 0.025$; PairedCre$^-$ versus UnpairedCre$^+$, $P = 0.706$; $^*P < 0.05$). Data are presented as means and error bars represent s.e.m.

unreinforced session. PairedCre$^+$ subjects responded more strongly to the visual cue on the first test trial than subjects from either control group (**Fig. 2e**,**f**), indicating greater learning. A two-way repeated-measures ANOVA revealed a significant interaction between group and trial ($F_{4,50} = 3.819$, $P = 0.009$) and a trend toward a main effect of group ($F_{2,25} = 3.272$, $P = 0.055$). Planned *post hoc* comparisons showed a significant difference between the PairedCre$^+$ group and PairedCre$^-$ ($P = 0.005$) or UnpairedCre$^+$ ($P < 0.001$) controls on the first test trial, whereas control groups did not differ (UnpairedCre$^+$ versus PairedCre$^-$, $P = 0.155$; **Fig. 2e**,**f**). This result indicates that unilateral VTA dopamine neuron activation at the time of unconditioned stimulus delivery was sufficient to cause new learning about preceding environmental cues. The observed dopamine neuron–induced learning enhancement was temporally specific, as responding to the visual cue was blocked in the UnpairedCre$^+$ group receiving optical stimulation outside of the cue and unconditioned stimulus periods. Notably, PairedCre$^+$ and UnpairedCre$^+$ rats received equivalent stimulation, and this stimulation was equally reinforcing (**Supplementary Fig. 2a**–**c**), so discrepancies in the efficacy of optical stimulation between the PairedCre$^+$ and UnpairedCre$^+$ groups cannot explain the observed behavioral differences.

One possible explanation for the behavioral changes that we observed in the blocking experiment is that optical stimulation of dopamine neurons during compound training served to increase the value of the paired sucrose reward. Such an increase in value would result in a RPE (although not encoded by dopamine neurons) and unblock learning. We found, however, that the manipulation of dopamine neuron activity during the consumption of one of two equally preferred, distinctly flavored sucrose solutions did not change the relative value of these rewards (measured as reward preference; Online Methods and **Supplementary Figs. 3** and **4**). This suggests that the unblocked learning about the newly added cue X was not the result of increased reward value induced by manipulating dopamine neuron activity.

Dopamine neuron activation slows extinction

Negative prediction errors also drive learned behavioral changes. For example, after a cue-reward association has been learned, decrementing or omitting the expected reward results in decreased reward-seeking behavior. Dopamine neurons show a characteristic pause in firing in response to reward decrements or omissions[8–10], and this pause is proposed to contribute to decreased behavioral responding to cues after reward decrement[4,29]. Having established that optogenetically activating dopamine neurons can drive new learning about cues under conditions in which dopamine neurons normally do not change their firing patterns from baseline levels, we next tested whether similar artificial activation at a time when dopamine neurons normally decrease firing could counter decrements in behavioral performance associated with reducing the value of the unconditioned stimulus. *Th-cre*$^+$ and *Th-cre*$^-$ rats that received unilateral ChR2-containing virus infusions and optical fiber implants targeted to the VTA (**Supplementary Fig. 1**) were trained to respond for sucrose whose availability was predicted by an auditory cue. The auditory cue was presented 1 d after the last training session, but water was substituted for the sucrose unconditioned stimulus (downshift test; **Fig. 3a**). PairedCre$^+$ and PairedCre$^-$ rats received dopamine neuron optical stimulation (3-s train, 5-ms pulse, 20 Hz) concurrent with water delivery when they entered the reward port during the cue; UnpairedCre$^+$ rats received stimulation during the ITI. Rats were subjected to a downshift recall session later; the recall session was identical to the initial extinction test,

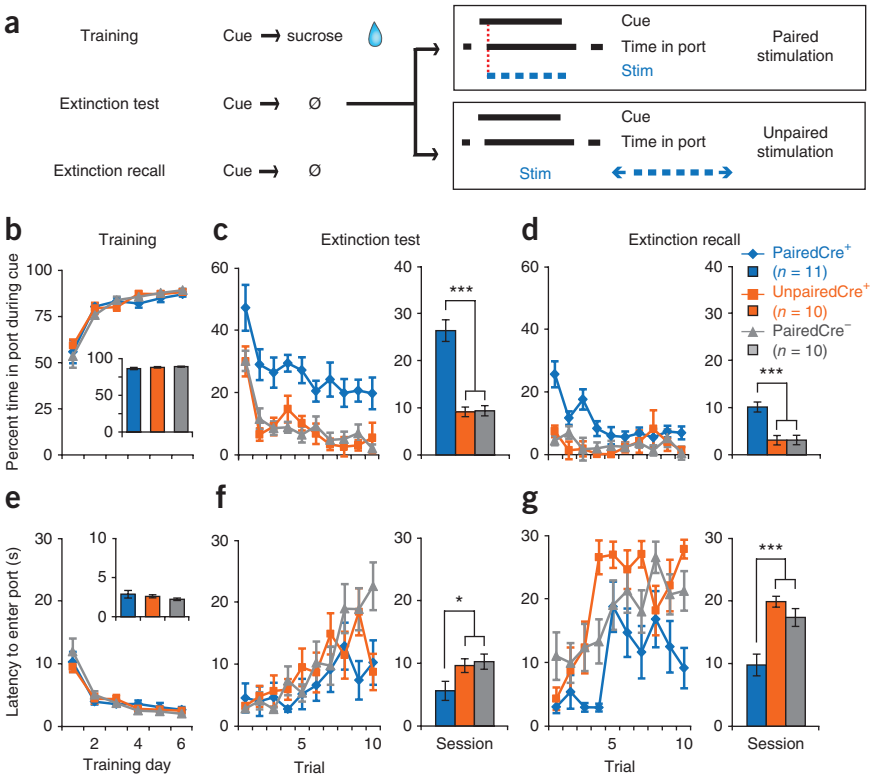

Figure 4 Dopamine neuron stimulation attenuates behavioral decrements associated with reward omission. (**a**) Experimental design for extinction experiment. Note that the same subjects from the downshift experiment were used for this procedure, with Cre$^+$ groups shuffled between experiments (see Online Methods). Optical stimulation (3-s train, 5-ms pulse, 20 Hz, 473 nm) was delivered at the time of expected reward for Paired groups and during ITI for UnpairedCre$^+$ rats during the extinction test. (**b**) Percent time in port during the cue across training sessions. Inset, no difference was observed in average performance during the last two training sessions. (**c**) Percent time in port during the cue for the extinction test. Data are displayed for single trials (left) and as a session average (right). PairedCre$^+$ rats exhibited increased time in port compared with controls (PairedCre$^+$ versus UnpairedCre$^+$, $^{***}P < 0.001$; PairedCre$^+$ versus PairedCre$^-$, $^{***}P < 0.001$; PairedCre$^-$ versus UnpairedCre$^+$, $P = 0.920$). (**d**) Percent time in port during the cue for extinction recall. Data are displayed for single trials (left) and as a session average (right). PairedCre$^+$ rats exhibited increased time in port compared with controls (PairedCre$^+$ versus UnpairedCre$^+$, $^{***}P < 0.001$; PairedCre$^+$ versus PairedCre$^-$, $^{***}P < 0.001$; PairedCre$^-$ versus UnpairedCre$^+$, $P = 0.984$). (**e**) Latency to enter the reward port after cue onset. Inset, no group differences were observed during the last two training sessions. (**f**) Data are presented as in **c**, but for latency. PairedCre$^+$ rats responded faster to the cue than controls during the extinction test (PairedCre$^+$ versus UnpairedCre$^+$, $P = 0.038$; PairedCre$^+$ versus PairedCre$^-$, $P = 0.04$; PairedCre$^-$ versus UnpairedCre$^+$, $P = 0.727$; $^*P < 0.05$). (**g**) Data are presented as in **d**, but for latency. PairedCre$^+$ rats responded faster to the cue than controls during extinction recall (PairedCre$^+$ versus UnpairedCre$^+$, $^{***}P < 0.001$; PairedCre$^+$ versus PairedCre$^-$, $^{***}P < 0.001$; PairedCre$^-$ versus UnpairedCre$^+$, $P = 0.211$). Data are presented as means and error bars represent s.e.m.

except that no optical stimulation was given. The purpose of the recall session was to determine whether optical stimulation had caused long-lasting behavioral changes. Cue responding was measured as the percent time spent in the reward port during the cue normalized to a pre-cue baseline (**Fig. 3b–d**) and as the latency to enter the reward port after cue onset (**Fig. 3e–g**).

All groups acquired the initial cue–reward association (**Fig. 3b**,**e**); a two-way repeated-measures ANOVA revealed no significant effects of group or group × day interactions at the end of training (all P values > 0.277). During the downshift test, PairedCre$^-$ and UnpairedCre$^+$ group performance rapidly deteriorated. This was evident on a trial-by-trial basis (**Fig. 3c**,**f**) and when cue responding was averaged across the entire downshift test session (**Fig. 3c**,**f**). In contrast, PairedCre$^+$ rats receiving optical stimulation concurrent with water delivery showed much reduced (**Fig. 3c**) or no (**Fig. 3f**) decrement in behavioral responding. Two-way repeated-measures ANOVAs revealed significant effects of group and group × trial interactions for both time spent in the port during the cue (group, $F_{2,28} = 11.12$, $P < 0.001$; group × trial, $F_{18,252} = 1.953$, $P = 0.013$) and latency to respond after cue onset (group, $F_{2,28} = 12.463$, $P < 0.001$; group × trial, $F_{18,252} = 4.394$, $P < 0.001$). Planned *post hoc* comparisons revealed that PairedCre$^+$ rats differed significantly from controls in both time and latency ($P < 0.001$), whereas control groups did not differ from each other ($P > 0.375$). Notably, some group differences persisted into the downshift recall session in which no stimulation was delivered (latency: main effect of group, $F_{2,28} = 4.597$, $P = 0.019$; **Fig. 3g**). These data indicate that phasic VTA dopamine neuron activation can partially counteract performance changes associated with reducing reward value.

We next examined whether our optical manipulation would be effective if the expected reinforcer was omitted entirely (**Fig. 4**). Rats used in the downshift experiment (see Online Methods) were trained on a new cue-reward association (**Fig. 4a**). All rats learned the new association (**Fig. 4b**,**e**); a two-way repeated-measures ANOVA revealed no significant effects of group or group × day interactions at the end of training (all P values > 0.242). Subsequently, all rats were subjected to an extinction test in which the expected sucrose reward was withheld. Instead, PairedCre$^+$ and PairedCre$^-$ rats received optical stimulation (3-s train, 5-ms pulse, 20 Hz) of dopamine neurons at the time of expected unconditioned stimulus delivery, whereas UnpairedCre$^+$ rats received optical stimulation during the ITI. Rats were subjected to an extinction recall session 1 d later in which neither the unconditioned stimulus nor optical stimulation were delivered to determine whether prior optical stimulation results in long-lasting behavioral changes.

During the extinction test, PairedCre$^+$ rats spent more time in the reward port during the cue and responded to the cue more quickly than both PairedCre$^-$ and UnpairedCre$^+$ rats (**Fig. 4c**,**f**); two-way repeated-measures ANOVAs revealed significant effects of group and/or group × trial interactions for both measures (percent time: group, $F_{2,28} = 40.054$, $P < 0.001$; group × trial, $F_{18,252} = 0.419$, $P = 0.983$; latency: group, $F_{2,28} = 3.827$, $P = 0.034$; group × trial, $F_{18,252} = 2.047$, $P = 0.008$), and these behavioral differences persisted into the extinction recall session (two-way repeated-measures ANOVAs, significant main effects of group and group × trial interactions, $F > 2$, $P < 0.01$ in all cases; **Fig. 4d**,**g**). Thus, VTA dopamine neuron activation at the time of expected reward is sufficient to sustain conditioned behavioral responding when expected reward is omitted. For both reward downshift and omission, the behavioral

effects of dopamine neuron stimulation were temporally specific, as UnpairedCre$^+$ rats responded less than PairedCre$^+$ rats despite receiving more stimulation during the test sessions (**Supplementary Fig. 2d**,**g**) and despite verification that this stimulation is equally reinforcing in both Cre$^+$ groups (**Supplementary Fig. 2e**,**f**,**h**,**i**).

Despite causing substantial behavioral changes during extinction, optogenetic activation of dopamine neurons failed to maintain reward-seeking behavior at pre-extinction levels. This may be a result of the inability of our dopamine neuron stimulation to fully counter the expected decrease in dopamine neuron firing during reward omission or downshift. Alternatively, this may reflect competition between the artificially imposed dopamine signal and other neural circuits specialized to inhibit conditioned responding when this behavior is no longer advantageous, as has been proposed[31,32].

Estrus cycle can modulate dopaminergic transmission under some circumstances[33]. Notably, although female rats were used in these studies, we tracked estrus stage during a behavioral session in which dopamine neurons were stimulated, and we failed to observe correlations between estrus cycle stage and behavioral performance (**Supplementary Fig. 5**).

DISCUSSION

We found that RPE signaling by dopamine neurons is causally related to cue-reward learning. We leveraged the temporal precision afforded by optogenetic tools to mimic endogenous RPE signaling in VTA dopamine neurons to examine how these artificial signals affect subsequent behavior. Using an associative blocking procedure, we observed that increasing dopamine neuron activity during reward delivery could drive new learning about antecedent cues that would not normally guide behavior. Using extinction procedures, we observed that reductions in conditioned responding that normally accompany decreases in reward value are attenuated when dopamine neuron activity is increased at the time of expected reward. Notably, the behavioral changes we observed in all experiments were long lasting, persisting 24 h after dopamine neurons were optogenetically activated, and temporally specific, failing to occur if dopamine neurons were activated at times outside of the reward consumption period. Taken together, our results indicate that RPE signaling by dopamine neurons is sufficient to support new cue-reward learning and modify previously learned cue-reward associations.

Our results clearly establish that artificially activating VTA dopamine neurons at the time that a natural reward is delivered (or expected) supports cue-elicited responding. A question of fundamental importance is why this occurs. In particular, for the blocking study, one possibility is that dopamine stimulation acted as an independent reward, discriminable from the paired sucrose reward, which initiated the formation of a parallel association between the reward-predictive cue and dopamine stimulation itself. However, this explanation assumes cue independence and would require the rat to compute two simultaneous, yet separate, prediction errors controlling the strength of two separate associations (cue A $\rightarrow$ sucrose, cue X $\rightarrow$ dopamine stimulation). Indeed, the assumption of cue independence was challenged[1] specifically because separate prediction errors cannot account for the phenomenon of blocking. If each cue generated its own independent prediction error, then the preconditioning of one cue would not affect the future conditioning of other cues, but it does, as the blocking procedure revealed. Blocking showed that cues presented simultaneously interact and compete for associative strength. Thus, it is unlikely that a parallel association formed between reward-predictive cues and dopamine stimulation can account for our results. Of interest, putative dopamine neurons do not appear to encode a sensory representation of reward, as they do not discriminate among rewards on the basis of their sensory properties[29]; thus, it is not obvious how dopamine neuron activation coincident with natural reward delivery could be perceived as distinct from that reward.

Although previous studies have suggested otherwise[34,35], another related possibility is that optical activation of dopamine neurons induces behavioral changes by directly enhancing the value of the paired natural reward. To address this possibility, we conducted a control study based on the idea that high-value rewards are preferred over less valuable alternatives. We paired dopamine neuron stimulation with consumption of a flavored, and therefore discriminable, sucrose solution; we reasoned that if dopamine neuron stimulation served to increase the value of a paired reward, this should manifest as an increased preference for the stimulation-paired reward over a distinctly flavored, but otherwise identical sucrose solution. However, we observed that reward that was previously paired with dopamine neuron stimulation was preferred equivalently to one that was not. This result does not support the interpretation that optical dopamine neuron stimulation supported learning in our experiments by increasing the value of the sucrose reward.

Alternatively, the behavioral changes that we observed in PairedCre$^+$ rats could reflect the development of a conditioned place preference for the location at which optical stimulation was delivered (that is, the reward port), as has been demonstrated[20]. If this were the case, we should have observed generalized increases in reward-seeking behavior across the entire test session. Notably, our primary behavioral metric (time spent in the reward port during the cue) was normalized to pre-cue baseline levels. If optical stimulation had induced nonspecific increases in reward-seeking behavior, our normalized measure should have approached zero. However, we found that reward-seeking was specifically elevated during cue presentation. Although we observed robust group differences in our normalized measures, a separate analysis of the absolute percent time spent in the port in the pre-cue baseline period during any test session revealed no significant group differences (all P values > 0.17). Together, these findings indicate that the behavioral changes that we observed are unlikely to be the result of a conditioned place preference.

Instead, the most parsimonious explanation for our results is that dopamine stimulation reproduced a RPE. Theories of associative learning hold that simple pairing, or contiguity, between a stimulus and reward or punishment is not sufficient for conditioning to occur; learning requires the subject to detect a discrepancy or prediction error between the expected and actual outcome that serves to correct future predictions[1]. Although compelling correlative evidence suggests that dopamine neurons are well-suited to provide such a teaching signal, little proof exists to support this notion. For this reason, our results represent an advance over previous work. Although prior studies that also used optogenetic tools to permit temporally precise control of dopamine neuron activity found that dopamine neuron activation is reinforcing, these studies did not establish the means by which this stimulation can reinforce behavior. Because we used behavioral procedures in which learning is driven by reward prediction errors, our data establish the critical behavioral mechanism (RPE) through which phasic dopamine signals timed with reward cause learning.

Through which cellular and circuit mechanisms could this dopamine signal cause learning to occur? Although few in number, VTA dopamine neurons send extensive projections to a variety of cortical and subcortical areas and are therefore well-positioned to influence neuronal computation[2,36–38]. Increases in dopamine neuron firing during unexpected reward could function as a teaching signal

used broadly in efferent targets to strengthen neural representations that facilitate reward receipt[39,40], possibly via alterations in the strength and direction of synaptic plasticity[37,41–43]. Because our artificial manipulation of dopamine neuron activity produced behavioral changes that lasted at least 24 h after the stimulation ended, such dopamine-induced, downstream changes in synaptic function may have occurred; in addition, both natural cue-reward learning[44] and optogenetic stimulation of dopamine neurons[45] alter glutamatergic synaptic strength onto dopamine neurons themselves, providing another possible basis for the long-lasting effects of dopamine neuron activation on behavior. One or both of these synaptic mechanisms may underlie the behavioral changes reported here. Although the physiological consequences of optogenetic dopamine neuron activation have been investigated in *in vitro* preparations and in anesthetized rats, to fully explore these synaptic mechanisms, a first critical step is to define the effects of optical activation on neuronal firing and on dopamine release in the awake behaving subject.

We focused on the role of dopamine neuron activation at the time of reward. Another hallmark feature of dopamine neuron firing during associative learning is the gradual transfer of neural activation from reward delivery to cue onset. Early in learning, when cue-reward associations are weak, dopamine neurons respond robustly to the occurrence of reward and weakly to reward-predictive cues. As learning progresses neural responses to the cue become more pronounced and reward responses diminish[10]. Although our results support the idea that reward-evoked dopamine neuron activity drives conditioned behavioral responding to cues, the function(s) of cue-evoked dopamine neuron activity remain a fruitful avenue for investigation.

Possible answers to this question have already been proposed. This transfer of the dopamine teaching signal from the primary reinforcer to the preceding cue is predicted by temporal-difference models of learning[46]. In such models, the back-propagation of the teaching signal allows the earliest predictor of the reward to be identified, thereby delineating the chain of events leading to reward delivery[2,6,46]. Alternatively, or in addition, cue-evoked dopamine may encode the cue's incentive value, endowing the cue itself with some of the motivational properties originally elicited by the reward, thereby making the cue desirable in its own right[34]. Using behavioral procedures that allow a cue's predictive and incentive properties to be assessed separately, a recent study provided evidence for dopamine's role in the acquisition of cue-reward learning for the latter, but not the former, process[47]. Such behavioral procedures could also prove useful to determine in greater detail how learning induced by mimicking RPE signals affects cue-induced conditioned responding. These and other future attempts to define the precise behavioral consequences of dopamine neuron activity during cues and rewards will further refine our conceptions of the role of dopamine RPE signals in associative learning.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We are grateful for the technical assistance of M. Olsman, L. Sahuque and R. Reese, and for critical feedback from H. Fields during manuscript preparation. This research was supported by the US National Institutes of Health (grants DA015096 and AA17072 to P.H.J. and a New Innovator award to I.B.W.), funds from the State of California for medical research on alcohol and substance abuse through the University of California, San Francisco (P.H.J.), a National Science Foundation Graduate Research Fellowship (E.E.S.), and grants from the National Institute of Mental Health, the National Institute on Drug Abuse, the Michael J Fox Foundation, the Howard Hughes Medical Institute, and the Defense Advanced Research Projects Agency Reorganization and Plasticity to Accelerate Injury Recovery Program (K.D.).

AUTHOR CONTRIBUTIONS

E.E.S., R.K. and P.H.J. designed the experiments. E.E.S., R.K. and J.R.B. performed the experiments. I.B.W. and K.D. contributed reagents. E.E.S., R.K. and P.H.J. wrote the paper with comments from all of the authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Rescorla, R.A. & Wagner, A.R. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. in *Classical Conditioning II: Current Research and Theory* (eds. Black, A.H. & Prokasy, W.F.) 64–99 (Appleton Century Crofts, New York, 1972).
2. Glimcher, P.W. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proc. Natl. Acad. Sci. USA* **108** (suppl. 3): 15647–15654 (2011).
3. Montague, P.R., Dayan, P. & Sejnowski, T.J. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16**, 1936–1947 (1996).
4. Schultz, W., Dayan, P. & Montague, P.R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
5. Schultz, W. & Dickinson, A. Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* **23**, 473–500 (2000).
6. Sutton, R.S. & Barto, A.G. Toward a modern theory of adaptive networks: expectation and prediction. *Psychol. Rev.* **88**, 135–170 (1981).
7. Schultz, W., Apicella, P. & Ljungberg, T. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J. Neurosci.* **13**, 900–913 (1993).
8. Cohen, J.Y., Haesler, S., Vong, L., Lowell, B.B. & Uchida, N. Neuron type–specific signals for reward and punishment in the ventral tegmental area. *Nature* **482**, 85–88 (2012).
9. Roesch, M.R., Calu, D.J. & Schoenbaum, G. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat. Neurosci.* **10**, 1615–1624 (2007).
10. Hollerman, J.R. & Schultz, W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat. Neurosci.* **1**, 304–309 (1998).
11. Mirenowicz, J. & Schultz, W. Importance of unpredictability for reward responses in primate dopamine neurons. *J. Neurophysiol.* **72**, 1024–1027 (1994).
12. Day, J.J., Roitman, M.F., Wightman, R.M. & Carelli, R.M. Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens. *Nat. Neurosci.* **10**, 1020–1028 (2007).
13. Takahashi, Y.K. *et al.* The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. *Neuron* **62**, 269–280 (2009).
14. Iordanova, M.D., Westbrook, R.F. & Killcross, A.S. Dopamine activity in the nucleus accumbens modulates blocking in fear conditioning. *Eur. J. Neurosci.* **24**, 3265–3270 (2006).
15. O'Tuathaigh, C.M. *et al.* The effect of amphetamine on Kamin blocking and overshadowing. *Behav. Pharmacol.* **14**, 315–322 (2003).
16. Parker, J.G. *et al.* Absence of NMDA receptors in dopamine neurons attenuates dopamine release, but not conditioned approach, during Pavlovian conditioning. *Proc. Natl. Acad. Sci. USA* **107**, 13491–13496 (2010).
17. Zweifel, L.S. *et al.* Disruption of NMDAR-dependent burst firing by dopamine neurons provides selective assessment of phasic dopamine-dependent behavior. *Proc. Natl. Acad. Sci. USA* **106**, 7281–7288 (2009).
18. Adamantidis, A.R. *et al.* Optogenetic interrogation of dopaminergic modulation of the multiple phases of reward-seeking behavior. *J. Neurosci.* **31**, 10829–10835 (2011).
19. Domingos, A.I. *et al.* Leptin regulates the reward value of nutrient. *Nat. Neurosci.* **14**, 1562–1568 (2011).
20. Tsai, H.C. *et al.* Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science* **324**, 1080–1084 (2009).
21. Witten, I.B. *et al.* Recombinase-driver rat lines: tools, techniques, and optogenetic application to dopamine-mediated reinforcement. *Neuron* **72**, 721–733 (2011).
22. Boyden, E.S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **8**, 1263–1268 (2005).
23. Zhang, F., Wang, L.P., Boyden, E.S. & Deisseroth, K. Channelrhodopsin-2 and optical control of excitable cells. *Nat. Methods* **3**, 785–792 (2006).
24. Kamin, L.J. "Attention-like" processes in classical conditioning. in *Miami Symposium on the Prediction of Behavior, 1967: Aversive Stimulation* (ed. Jones, M.R.) 9–31 (University of Miami Press, 1968).

25. Kamin, L.J. Selective association and conditioning. in *Fundamental Issues in Associative Learning* (eds. Mackintosh, N.J. & Honig, F.W.K.) 42–64 (Dalhousie University Press, 1969).
26. Kamin, L.J. Predictability, surprise, attention and conditioning. in *Punishment and Aversive Behavior* (eds. Campbell, B.A. & Church, R.M.) 279–296 (Appleton-Century-Crofts, New York, NY, 1969).
27. Holland, P.C. Unblocking in Pavlovian appetitive conditioning. *J. Exp. Psychol. Anim. Behav. Process.* **10**, 476–497 (1984).
28. Waelti, P., Dickinson, A. & Schultz, W. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* **412**, 43–48 (2001).
29. Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80**, 1–27 (1998).
30. Burke, K.A., Franz, T.M., Miller, D.N. & Schoenbaum, G. The role of the orbitofrontal cortex in the pursuit of happiness and more specific rewards. *Nature* **454**, 340–344 (2008).
31. Daw, N.D., Kakade, S. & Dayan, P. Opponent interactions between serotonin and dopamine. *Neural Netw.* **15**, 603–616 (2002).
32. Peters, J., Kalivas, P.W. & Quirk, G.J. Extinction circuits for fear and addiction overlap in prefrontal cortex. *Learn. Mem.* **16**, 279–288 (2009).
33. Becker, J.B. Gender differences in dopaminergic function in striatum and nucleus accumbens. *Pharmacol. Biochem. Behav.* **64**, 803–812 (1999).
34. Berridge, K.C. & Robinson, T.E. What is the role of dopamine in reward: hedonic impact, reward learning or incentive salience? *Brain Res. Rev.* **28**, 309–369 (1998).
35. Wassum, K.M., Ostlund, S.B., Balleine, B.W. & Maidment, N.T. Differential dependence of Pavlovian incentive motivation and instrumental incentive learning processes on dopamine signaling. *Learn. Mem.* **18**, 475–483 (2011).
36. Beckstead, R.M., Domesick, V.B. & Nauta, W.J. Efferent connections of the substantia nigra and ventral tegmental area in the rat. *Brain Res.* **175**, 191–217 (1979).
37. Fields, H.L., Hjelmstad, G.O., Margolis, E.B. & Nicola, S.M. Ventral tegmental area neurons in learned appetitive behavior and positive reinforcement. *Annu. Rev. Neurosci.* **30**, 289–316 (2007).
38. Swanson, L.W. The projections of the ventral tegmental area and adjacent regions: a combined fluorescent retrograde tracer and immunofluorescence study in the rat. *Brain Res. Bull.* **9**, 321–353 (1982).
39. Reynolds, J.N. & Wickens, J.R. Dopamine-dependent plasticity of corticostriatal synapses. *Neural Netw.* **15**, 507–521 (2002).
40. Wickens, J.R., Horvitz, J.C., Costa, R.M. & Killcross, S. Dopaminergic mechanisms in actions and habits. *J. Neurosci.* **27**, 8181–8183 (2007).
41. Gerfen, C.R. & Surmeier, D.J. Modulation of striatal projection systems by dopamine. *Annu. Rev. Neurosci.* **34**, 441–466 (2011).
42. Reynolds, J.N., Hyland, B.I. & Wickens, J.R. A cellular mechanism of reward-related learning. *Nature* **413**, 67–70 (2001).
43. Tye, K.M. *et al.* Methylphenidate facilitates learning-induced amygdala plasticity. *Nat. Neurosci.* **13**, 475–481 (2010).
44. Stuber, G.D. *et al.* Reward-predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. *Science* **321**, 1690–1692 (2008).
45. Brown, M.T. *et al.* Drug-driven AMPA receptor redistribution mimicked by selective dopamine neuron stimulation. *PLoS ONE* **5**, e15870 (2010).
46. Suri, R.E. TD models of reward predictive responses in dopamine neurons. *Neural Netw.* **15**, 523–533 (2002).
47. Flagel, S.B. *et al.* A selective role for dopamine in stimulus-reward learning. *Nature* **469**, 53–57 (2011).

ONLINE METHODS

Subjects and surgery. 115 female transgenic rats (Long-Evans background) were used in these studies; 68 rats expressed Cre recombinase under the control of the *tyrosine hydroxylase* promoter (*Th-cre*$^+$) and 47 rats were their wild-type littermates (*Th-cre*$^-$). All rats weighed >225 g at the time of surgery. During testing (except the flavor preference study), rats were mildly food restricted to 18 g of lab chow per day given after the conclusion of daily behavioral sessions; on average, rats maintained >95% free-feeding weight. Water was available *ad libitum* in the home cage. Rats were singly housed under a 12-h:12-h light/dark cycle, with lights on at 7 a.m. The majority of behavioral experiments were conducted during the light cycle. Animal care and all experimental procedures were in accordance with guidelines from the US National Institutes of Health and were approved in advance by the Gallo Center Institutional Animal Care and Use Committee. We used stereotaxic surgical procedures for VTA infusion of Cre-dependent virus (Ef1α-DIO-ChR2-eYFP)[20] and optical fiber placement as previously described[21], with the exception that dorsoventral coordinates were adjusted to account for the smaller size of female rats as follows: dorsoventral –8.1 and –7.1 mm below skull surface for virus infusions and –7.1 mm for optical fiber implants.

Behavioral procedures. All behavioral experiments were conducted >2 weeks post-surgery; sessions that included optical stimulation were conducted >4 weeks post-surgery.

Apparatus. Behavioral sessions were conducted in sound-attenuated conditioning chambers (Med Associates). The left and right walls were fitted with reward delivery ports; computer-controlled syringe pumps located outside of the sound-attenuating cubicle delivered sucrose solution or water to these ports. The left wall had two nosepoke ports flanking the central reward delivery port; each nosepoke port had three LED lights at the rear. Chambers were outfitted with 2,700-Hz pure tone and white noise auditory stimuli, both delivered at 70 dB, as well as a 28-V chamber light above the left reward port. During behavioral sessions, the pure tone was pulsed at 3 Hz (0.1 s on, 0.2 s off) to create a stimulus that was easily distinguished from continuous white noise.

Reward delivery. All experiments (except the flavor preference study) involved delivery of a liquid sucrose solution (15%, wt/vol) during the presentation of auditory or combined auditory-visual cues. During each cue, entry into the active port triggered a 3-s delivery of sucrose solution (0.1 ml). After a 2-s timeout, another entry into the port (or the rat's continued presence at the port) triggered an additional 3-s reward delivery. This 5-s cycle could be repeated up to six times per 30-s trial, depending on the rat's behavior. For sessions in which optical stimulation was delivered, the laser was activated each time sucrose was delivered (or expected; **Figs. 2c**, **3a** and **4a**). This method of reward delivery, where reward and optical stimulation were both contingent on the rat's presence in the active port, was used for all experiments as it allowed for the coincident delivery of natural rewards and optical stimulation and maximized the temporal precision of reward expectation.

Blocking procedure. Rats received a 1-d habituation session where all auditory and visual cues used during future training sessions, as well as the sucrose reward, were presented individually (three presentations of each cue, 5-min ITI; ~60 reward deliveries, 1-min ITI). This session was intended to minimize unconditioned responses to novel stimuli and shape reward-seeking behavior to the correct (left) reward port. Next, rats underwent single-cue training where one of two auditory cues (white noise or pulsed tone, counterbalanced across subjects) was presented for 30 s on a variable interval, 4-min schedule for ten trials per session. Sucrose was delivered during each cue as described above. After 14–15 sessions of single cue training, compound cue training commenced and lasted for 4 d. During this phase, either the same auditory cue used in single cue training (blocking groups) or a new auditory cue (control group) was presented simultaneously with a visual cue. The visual cue consisted of the chamber light, which was the sole source of chamber illumination, flashing on/off at 0.3 Hz (1 s on, 2 s off). Sucrose reward was delivered as described above during this phase. A probe test was administered 24 h after the conclusion of the last compound training session to assess conditioned responding to the visual cue. During this session the visual cue was presented alone in the absence of sucrose, auditory cues or optical stimulation.

Downshift procedure. Rats received one session where sucrose reward was delivered to the active (right) port (50 deliveries, 30 s, variable interval) to shape reward-seeking behavior to this location. Subsequently, rats were trained to respond for sucrose during an auditory cue (white noise) as described above in 11 daily sessions. A downshift test session was administered 24 h later that was identical to previous training sessions except that water was substituted for sucrose and optical stimulation was delivered coincidently. A downshift recall test was administered 24 h later, in which water was delivered during the cue, but optical stimulation did not occur.

Extinction procedure. This experiment was conducted 2 weeks after the end of the downshift experiment with the same subjects; group assignment for Cre$^+$ rats was shuffled between experiments. Rats received one session of sucrose reward delivery to the opposite (left) port used in the downshift test to shape reward-seeking behavior to this location. Subsequently, rats were trained to respond for sucrose during an auditory cue (pulsed tone) as described above in six daily training sessions. An extinction test session was administered 24 h later that was identical to previous training sessions except that no reward was given and optical stimulation was delivered at the time that the sucrose reward had been available in previous training. 24 h later, an extinction recall test was administered in which the auditory cue was presented, but no reward or optical stimulation was delivered.

Intracranial self-stimulation (ICSS). Following completion of the experiments described above, all rats were given four daily 1-h sessions of ICSS training, as described previously[21]. Food restriction ceased at least 24 h before the first ICSS session. A response at the nosepoke port designated as active resulted in the delivery of a train of light pulses matched to the stimulation parameters used in that subject's previous behavioral experiment (1 s, 20 Hz for rats in blocking or flavor preference studies, 3 s, 20 Hz for rats in downshift or extinction studies).

Flavor preference study. Rats were initially trained to drink unflavored 15% sucrose solution (wt/vol) from the reward port in the conditioning chambers (0.1 ml delivered on variable interval 30-s schedule, 50 deliveries). Rats were then given overnight access to 40 ml each of the flavored sucrose solutions (15% sucrose + 0.15% Kool-Aid, tropical punch or grape flavors, wt/vol) in their home cage to ensure that all subjects had sampled both flavors before critical consumption tests. Home cages were equipped with two bottle slots; before the start of the experiment both slots were occupied by water bottles to reduce possible side bias.

For home cage consumption tests, water bottles were removed from the home cage 15–30 min before the start of consumption tests. A standardized procedure was used to ensure that rats briefly sampled both flavors before free access to the solutions began. The purpose of this procedure was to make sure that rats were aware that both flavors were available, so that any measured preference reflected true choice. The experimenter placed a flavor bottle on the left side of the cage until the rat consumed the solution for 2–3 s. This bottle was removed and a second bottle containing the other flavored solution was placed on the right side of the cage until the rat consumed the new solution for 2–3 s. The second bottle was then removed and both bottles were simultaneously placed on the home cage to start the test. The bottles were removed 10 min later and the amounts consumed were recorded. The cage side assigned to each flavor (left or right) alternated between consumption tests to control for possible side bias.

Flavor training began in the conditioning chambers 24 h after the home cage baseline consumption test (eight sessions total). Only one flavored sucrose solution was available per day; training days with each flavor were interleaved. One of the two flavors was randomly assigned for each rat to be the stimulated flavor. On training days when the stimulated flavor was available, optical stimulation was either paired with reward consumption for PairedCre$^+$ and PairedCre$^-$ groups, or explicitly unpaired (presented during the ITI at times when no reward was available) for UnpairedCre$^+$ rats (**Supplementary Fig. 4**). Flavored sucrose was delivered to a reward port on a variable interval 30-s schedule, with the exception that each reward had to be consumed before the next would be delivered. A reward was considered to be consumed if the rat maintained presence in the port for 1s or longer. Sessions lasted until the maximum of 50 rewards were consumed or 1 h elapsed, whichever occurred first. The final home cage preference test was conducted 24 h after the last flavor training session.

Optical activation. Intracranial light delivery in behaving rats was achieved as described[21]. For all experiments, 5-ms light pulses were delivered with a 50-ms inter-pulse interval (that is, 20 Hz). For blocking and flavor preference experiments, 20 pulses were used (1 s of stimulation). For downshift and extinction experiments, 60 pulses were used (3 s). Data from sessions where light output was compromised because of broken or disconnected optical cables was discarded and these subjects were excluded from the study. This criterion led to the exclusion of one rat from each of the blocking and extinction experiments, and four rats from the self-stimulation protocol.

Assessment of estrus cycle. Stage of estrus cycle was assessed by vaginal cytological examination using well-established methods[48]. After behavioral sessions (downshift study), the tip of a moistened cotton swab was gently inserted into the exterior portion of the vaginal canal and then rotated to dislodge cells from the vaginal wall. The swab was immediately rolled onto a glass slide, and the sample preserved with spray fixative (Spray-Cyte, Fisher Scientific) without allowing the cells to dry. Samples were collected over five consecutive days to ensure observation of multiple estrus cycle stages. This was done to improve the accuracy of determining estrus cycle stage on any single day of the experiment. Slides were then stained with a modified Papanicolau staining procedure as follows: 50% ethyl alcohol, 3 min; tap water, 10 dips (×2); Gill's hematoxylin 1, 6 min; tap water, 10 dips (×2); Scott's water, 4 min; tap water, 10 dips (×2); 95% ethyl alcohol, 10 dips (×2); modified orange-greenish 6 (OG-6), 1 min; 95% ethyl alcohol, 10 dips; 95% ethyl alcohol 8 dips; 95% ethyl alcohol 6 dips; modified eosin azure 36 (EA-36), 20 min; 95% ethyl alcohol, 40 dips; 95% ethyl alcohol, 30 dips; 95% ethyl alcohol, 20 dips; 100% ethyl alcohol, 10 dips (×2); xylene, 10 dips (×2); coverslip immediately. All staining solutions (Gill's hematoxylin 1, OG-6, and EA-36) were sourced from Richard Allen Scientific. Estrus cycle stage was determined by identifying cellular morphology characteristic to each phase according to previously described criteria[48].

Histology. Immunohistochemical detection of YFP and tyrosine hydroxylase was performed as described previously[21]. Although optical fiber placements and virus expression varied slightly from subject to subject, no subject was excluded on the basis of histology (**Supplementary Fig. 1**).

Data analysis. Counterbalancing procedures were used to form experimental groups that were balanced in terms of age, weight, conditioning chamber used, cue identity and behavioral performance in the sessions preceding the experimental intervention. Conditioned responding was measured as the amount of time spent in the reward port during cue presentation, normalized by subtracting the time spent in the port during a pre-cue period of equal length. Note that, during reinforced training sessions, this measure is not a pure index of learning, as the time spent in the port during the cue also reflects time spent consuming sucrose. For the blocking experiment, we focused exclusively on this measure because it proved to be particularly robust. Notably, during the blocking test itself, this measure is a pure index of learning because no reward is delivered during this session. For other experiments, we also measured the latency to enter the reward port after cue onset. Pilot experiments and power analyses for both the blocking and the extinction study indicated that 8–10 subjects per group allowed for detection of differences between experimental and control conditions, with $\alpha = 0.05$ and $\beta = 0.80$. In cases in which behavioral data from individual subjects varied from the group mean by more than two s.d. (calculated with data from all subjects included), these subjects were excluded as statistical outliers (two rats from the blocking experiment and three each from the downshift and extinction experiments) and their data were not further analyzed. Behavioral measures were analyzed using a mixed factorial ANOVA with the between-subjects factor of experimental group and the within-subjects factor of session or trial, followed by planned Student Newman-Keul's tests when indicated by significant main effects or interactions. For all tests, $\alpha = 0.05$, and all statistical tests were two-sided. By their design, the experiments focused on three planned comparisons (PairedCre$^+$ versus UnpairedCre$^+$, PairedCre$^+$ versus PairedCre$^-$, UnpairedCre$^+$ versus PairedCre$^-$). We found no major deviation from the assumptions of the ANOVA. For the cases in which normality or equal variance was questionable, the results of the ANOVA were confirmed by non-parametric tests (Kruskal-Wallis followed by *post hoc* Dunn's test). Although explicit blinding procedures were not employed, experimental group allocation was not noted on subject cage cards, and all behavioral data were collected automatically via computer. Blocking and extinction experiments are replications of pilot experiments. Although based on a pilot study, the flavor preference study was conducted as described only one time, but with sufficient sample size to make statistical inferences.

48. Karim, B.O. *et al*. Estrous cycle and ovarian changes in a rat mammary carcinogenesis model after irradiation, tamoxifen chemoprevention and aging. *Comp. Med.* **53**, 532–538 (2003).

doi:10.1038/nn.3413